THE UNIVERSITY OF ALBERTA

MEASURING INVENTIVENESS IN SENIOR HIGH SCHOOL MATHEMATICS

by

©

J. MODUPE TAYLOR-PEARCE

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF SECONDARY EDUCATION

EDMONTON, ALBERTA

SPRING, 1971

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled "Measuring Inventiveness in Senior High School Mathematics," submitted by J. Modupe Taylor-Pearce in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

ABSTRACT

The purpose of the study was to establish principles for constructing and scoring tests of inventiveness in mathematics at the senior high school level. Theoretical and experimental methods were used to establish these principles.

The statement of J. P. Guilford that "most of the more obvious contributions to creative thinking are in the divergent production category," was interpreted as a proposition that divergent production leads to inventiveness. The principles formulated in the study for constructing and scoring tests of inventiveness in mathematics were based on the following three fundamental propositions:

(i) That it is possible to evoke and measure inventiveness in mathematics,

(ii) that divergent production in mathematics leads to inventiveness in mathematics, and

(iii) that individual differences of students in some aspects of inventiveness in mathematics may be ascertained from their individual differences in the components of divergent production in mathematics.

The experimental establishment of the principles consisted of validating tests of inventiveness in mathematics constructed and scored using the formulated principles for divergent-production mathematics tests. Divergent-production tests were administered to samples of senior high school students in a preliminary study and in two main studies. Fourteen divergent-production tests were administered to one or the other of the two main samples, with six

of these tests being common to both samples.

Analytical investigations were conducted on the content, construct, and criterion validities of the tests. The divergent-production tests were primarily designed to measure specific divergent production abilities. The investigation on content validity was conducted by obtaining the assessments of experts in mathematics, mathematics education, and measurement, on the face and content validities of the tests. The investigation on construct validity took the form of ascertaining to what extent certain vital expectations arising from the theoretical bases of the tests were experimentally met. The investigations on criterion validity related to the extent to which the tests evoked and measured inventiveness.

It was found that the experimental evidence supported the fundamental propositions of the study. All fourteen divergent-production tests of the main studies were found to be valid as tests of inventiveness in senior high school mathematics, although it was found that some had satisfied more critical expectations for validity than others. The empirical validation of the tests of the main study concluded the establishment of the principles of the study.

Some important aspects of the study were discussed. Comment was provided on the fundamental propositions of the study, the principles of the study, and on Guilford's structure-of-intellect theory as applied to divergent production in school mathematics. Observed and inferred characteristics of the tests were high-lighted, and uses of divergent-production approaches to achieve the objective of inventiveness in classroom mathematics were discussed. Some problems for further research were presented.

ACKNOWLEDGEMENTS

I am very grateful to the Canadian Commonwealth Scholarship and Fellowship Committee for awarding me a scholarship to pursue a Ph.D. degree. I am very grateful to the Ministry of Education, Sierra Leone, for granting me study leave to pursue this degree.

I am very grateful to my advisor, Dr. S. E. Sigurdson, for the encouragement and guidance that I received from him throughout the study. His keen interest in the study encouraged me greatly. His searching questions forced me to deeper thinking.

I am very grateful to Dr. T. E. Kieren for his help and encouragement in the study. His confidence in me, and his requests for improvement on various aspects of the study, urged me to greater effort.

I am very grateful to Dr. T. Maguire for his help and encouragement, especially in the psychometric aspects of the study. I thank Dr. D. Nelson for his keen interest in the study, and the encouragement he gave me.

I am very grateful to Dr. H. I. Freedman for his keen interest in the study, and the help and encouragement he gave me in various aspects of the study.

I am very grateful to the Edmonton Public and Separate School Boards for the permission granted me to conduct experiments in some of their schools. I am very grateful, indeed, to the students of these schools who participated in the experiment, and the members of the staff who gave me considerable assistance.

I thank Dr. E. S. Keeping, Dr. H. I. Freedman, Dr. S. Willard, Dr. T. Maguire, Dr. G. Cathcart, Dr. L. D. Nelson, Dr. T. E. Kieren, Dr. S. E. Sigurdson, Mr. C. Climenhaga, Mr. A. F. Labor, Dr. W. Liedtke, Mr. J. Isaac, Mr. B. G. TeKampe, Miss A. K. Sunde, and Mrs. O. Taylor-Pearce, for acting as judges in various aspects of the study.

I thank Mrs. O. Taylor-Pearce, Miss A. K. Sunde, and Mr. B. G. TeKampe, for help in re-marking the papers, and Mr. A. F. Labor for help in the scoring of multiple choice test papers.

I am very grateful to Mrs. Farrella Bickersteth for typing the thesis so earnestly, so carefully, and so well.

I am grateful to Mr. Sheka Kanu for our discussions which affected my writing of the thesis. I am grateful to Mr. Jonathan Peters and Miss Halia Boychuk for help in proof reading.

I am very grateful to my mother, Mrs. E. Taylor-Pearce for her continued encouragement throughout my academic career.

I am very grateful to my wife, Olive, and the children, Funlayo, Dare, and Shola, for their love. My wife, more than anyone else, knows the difficulties and triumphs that marked the course of my study.

I thank my Savior, Jesus Christ, for everything. He has kept His promise to be with me always.

CONTENTS

CHAPTER PAGE

LIST OF TABLES

CHAPTER I

INTRODUCTION TO THE PROBLEM

1.1 INTRODUCTION

One of the educational objectives in mathematics that the International Study of Achievement in Mathematics "believed would be accepted as desirable by most teachers of mathematics regardless of their nationality," was "Inventiveness: reasoning creatively in mathematics" (Husen, 1967, p. 81). There is a growing concern that education should result not only in students being able to reproduce content and ideas, but also in their being able to produce new and useful ideas. Creative production which may previously have been conceived as a prerogative of the exceptionally able, is now being considered by some researchers as within the reach of everyone, and capable of being facilitated through education. Various evidences point to a recent and current ferment of activity on research into creativity, and J. P. Guilford (1959), one of the foremost researchers in the field gives a number of possible reasons for this. Investigations on creative behavior have had considerable implications for education, and much research has been conducted with students as subjects. One of the most important reasons for fostering creativity in education is the strong possibility that creativity is closely connected with learning. Some would even consider creative behavior in a subject area as a necessary condition for understanding and mastering the subject.

Many writers have emphasised that some form of creative activity is essentially involved in learning, understanding and retention. Bloom's Taxonomy notes that "In one sense, all learning is creative, the individual has acquired an understanding or some other reorganization of experience which is novel for him. The novelty for him is what makes the experience "creative." (Bloom, 1965, p. 165). Henle (1962, p. 35), argues that "following another's thinking is novel for the learner himself ... the process is still new ...." The mathematician A. Ya. Kinchkin draws from his own experience to emphasize a relation between creative, active work, and meaningful learning:

> If we carefully analyse our own experience, we would all unanimously agree that only those scientific facts regularly remain firmly and actively rooted in our memories which at some time were the object or the tool of our own work, our own creative activity. A book or an article, even if it has been read three times through, will inevitably be forgotten if its material has only been absorbed passively, if its contents have never been the raw material or the tool of our own active, creative work. (A. Ya. Kinchkin, 1968, p. 72).

When behavioral objectives involving creative behavior have been classified among other objectives of instruction, there has usually been a suggestion that they are part of a hierarchical scale in which the objectives related to the reproduction of subject matter have been considered as involved in the "lower" end, and the objectives related to production and invention as involved in a "higher" end. Thus Bloom's Taxonomy (1956) consists of six classes -- Knowledge, Comprehension, Application, Analysis, Synthesis, and Evaluation, and these categories according to the compilers "appear to us to represent something of the hierarchical order of the different classes of

objectives." (Bloom, 1956, p. 18). Some behaviors were conceived as including others: "As we have defined them, the objectives in one class are likely to make use of and be built on the behaviors found in the preceding classes in this list." Bloom claims that the category of synthesis "is the category in the cognitive domain which most clearly provides for creative behavior on the part of the learner." Synthesis is defined in the Taxonomy as the "putting together of elements and parts so as to form a whole." Thus in conceiving of synthesis as its penultimate category, Bloom's Taxonomy is suggesting that the behaviors of the previous categories are likely to be utilized in synthesis.

The objectives used by the International Project of Educational Achievement (IEA) in their "International Study of Achievement in Mathematics," also group creative behavior as a higher mental process. Their five objectives are:

A: Knowledge and information: recall of definitions, notation, concepts.

B: Techniques and skills: solutions

C: Translation of data into symbols or schema and vice versa

D: Comprehension: Capacity to analyse problems, to follow reasoning.

E: Inventiveness: reasoning creatively in mathematics.

According to the classifiers, "behaviors A and B may be classified as "lower" mental processes, while behaviors D and E may be regarded as "higher" mental processes." (Husen, 1967, p. 82).

Wood (1968), in his "Item Bank Project," employs a classification of objectives which he states is a cross between the IEA

classes and Bloom's Taxonomy. His classification is as follows:

A: Knowledge and information: recall of definitions, notations, concepts.

B: Techniques and skill: computation, manipulation of symbols.

C: Comprehension: capacity to understand problems, to translate symbolic forms, to follow and extend reasoning.

D: Application: of appropriate concepts in unfamiliar mathematical situations.

E: Inventiveness: reasoning creatively in mathematics.

Wood used the above classification in connection with an analysis by six mathematics teachers of their instructional objectives. He rates behavior E as the highest level of behavior reviewed. He reports that the teachers generally considered _Inventiveness_ to be an experimental category -- beyond the reach of most students:

> In practice, _Inventiveness_ was rarely cited by the teachers who used the classification because it was generally felt that it was beyond the capacity of the majority of their pupils. It was regarded throughout as an experimental category and needs a lot more attention before it can be confidently used. (Wood, 1968, p. 92).

While _Inventiveness_ may be considered an experimental category, there are indications that inventiveness is within the capacity of the majority of students. Evans (1964) constructed tests to measure "the ability of students to respond in creative mathematical situations at the late elementary and junior high school level," and reports that _all_ the students at all the grade levels he investigated were able to make some responses:

> While it is true that certain of the students did better in the tests than most of the others, it was also found that _all_ the students at all four grade levels were able to make some responses. Thus everyone was able to make use of the introductory material and examples to generate new ideas, however basic and simple these ideas may have been. This suggests the possibility that the

> classroom teacher might provide experiences which enable all of his students, at their own level of development, to have a part in formulating mathematical concepts (Evans, 1964, p. 201).

The investigator (Taylor-Pearce, 1969) found a very similar situation among senior high school students, and concluded that one of the most noteworthy results of his tests was that it enabled the students to formulate mathematical concepts. The indications were that this type of testing could be used to much advantage in the teaching/learning situation. (Taylor-Pearce, 1969, p. 20). Bruner's conception of creativity seems to be in harmony with these findings and deductions:

> For at any level of energy or intelligence there can be more or less of creating in our sense. Stupid people create for each other as well as benefiting from what comes from afar. So too, do slothful and torpid people. I have been speaking of creativity, not genius. (Bruner, 1962, p. 17).

If inventiveness should be a <u>de facto</u> objective of instruction in mathematics, then serious attempts should be made to solve a number of problems which include:

(a) How can this behavior be identified?

(b) How can it be measured

(c) How relevant is instruction to inventiveness?

(d) How can the results of creativity research be used in the teaching/learning situation?

Considering inventiveness as an instructional objective implies that instruction is related to inventiveness. Clearly, one basic aim of instruction is to impart information, and many writers emphasize that information is a necessary, although certainly not a sufficient condition for creative production.

Guilford (1967b) notes that "although we may agree with Albert Einstein that imagination is more important than knowledge, we must admit that knowledge or information is a requisite for creative thinking" (p. 437). Davis (1966) reports a creative encounter in the classroom, and concludes that "the early introduction of important ideas, does not merely aid learning -- it also facilitates creativity." He reports that a third grade boy invented an algorithm for subtracting that was in many ways the best he had seen. He claims that the boy could not possibly have invented his algorithm for subtracting if he had not acquired a proficiency in the arithmetic of signed numbers (Davis, 1961, p. 360). Harding is quoted by Patrick as noting that:

> Before anyone can give himself up to inspiration he must have acquired a mastery over his subject in order that the technical aspects should be in no way a hindrance to him. (Patrick, 1955, p. 8).

Mordell (1959), reflects that his special knowledge of modular functions enabled him to prove without too much trouble a well-known conjecture by Ramanujan on the function which had seemed to others difficult to prove. (Mordel, 1959, p. 20).

The acquiring of information may be considered as an essential part of a first stage in the creative process which has often been identified as the stage of preparation, following Wallas (1926; 1945). Guilford (1964) considers that a creator's whole past life may contribute to prepration for any creative act. (p. 171). Patrick (1955) notes that Szekely states that there is a functional relationship between knowledge or previous experience and productive, creative thinking, (p. 6). Hadamard (1945) feels that it is obvious that invention or discovery, be it in mathematics or anywhere else, takes

place by combining ideas. (Hadamard, 1945, p. 29). It would seem logical to suggest that the basic ideas which are synthesized into creative productions consist of mastered information or knowledge.

Instruction in mathematics should add to the mathematical experience of the student, and thus provide an increase in the ideational units that could later be synthesized into new productions. Thus tests of inventiveness subsequent to instruction in mathematics, might well aim at giving students opportunities for utilising their previous mathematical experiences, and in particular the experiences connected with learning the mathematical content under consideration.

If instruction in mathematics should facilitate inventiveness, it would seem essential for the teaching/learning situation that the attainment of this objective should be measured. This measurement should provide the teacher with feedback which should help in teaching and planning, and should afford measures of the individual differences of students in terms of the attainment of this objective.

The present study is designed to develop principles for valid and reliable measurements of the inventiveness of students who are studying mathematics at the Senior High School Level.

1.2 PROBLEM

The study seeks to establish principles for constructing and scoring tests of inventiveness in mathematics at the senior high school level.

1.2-1 The Significance of the Problem

The study is expected to contribute towards making it more practicable for teachers to consider inventiveness as a *de facto* instructional objective and to furnish guidelines for measuring the

attainment of this objective. Plato has been quoted as noting that "what is honoured in a country will be cultivated there" (Torrance, 1965, p.1). One of the ways of giving honourable recognition to an objective in a school situation is to make it a criterion for evaluation. Taba (1962, p. 313) observes that what is tested considerably affects what is taught and learned.

> The way of evaluating what is learned dictates the way in which learning takes place. The scope of evaluation determines what types or levels of learning are emphasized, no matter what the curriculum indicates. Furthermore, no matter what the teacher stresses, the student will selectively address himself to that learning on which he is examined ...... If a thoughtful reorganization of knowledge is stressed in the classroom, but the testing and grading are confined to the mastery of facts, the latter learning is reinforced. If creativity and thinking are stressed in evaluating student progress, factual cramming is less likely to be the order of the day.

There is the weighty possibility that one of the best ways to ensure that creative behavior is encouraged and reinforced in schools is by having it as a measurable criterion of evaluation of the progress of pupils.

Parnes (1965) feels that educators have a challenge "to help each individual student to maximise his potential." Thus it is conceived to be essential to nurture the creative talents of a person in his move towards "maximum self-realization." This study is expected to make some contribution towards the nurture of creative talent in the classroom.

This study will also attempt to relate some of the conclusions of J. P. Guilford to mathematics. The need for relating his theories and conclusions in creativity research to mathematics education has been expressed, and a number of studies have been concerned with

aspects of the problem (as for example Praise, 1967). This study is a further contribution in the field.

The study is a further investigation into subject-related creativity testing. In a previous study (Taylor-Pearce, 1969), the investigator constructed subject-specific tests and drew certain conclusions for mathematics teaching and learning from the results of testing. Certain problems attendant on this type of testing were encountered and formulated in connection with the determination of the flexibility score. This study is expected to continue these investigations, and to make further recommendations for the teaching and learning of mathematics.

The study emphasizes the utilization of instructional content in testing for inventiveness in the classroom. Some studies (e.g. Canisti, 1962) attempt to reduce the amount of instructional content in tests relating to creativity or reasoning to a minimum. The goal of utilizing inventiveness as a *de facto* objective of instruction necessitates its measurement in the context of the curriculum content of mathematics. To this end, it seems justified to conjecture that the type of testing conceived and advocated in this study, will be of crucial importance.

CHAPTER II

REVIEW OF SOME RELATED LITERATURE

2.1 INTRODUCTION

The problem of the measurement of inventiveness in mathematics is treated in this study as part of the problem of the measurement of creativity in general. In this chapter, major ideas relating to the identification and measurement of creativity are reviewed, and propositions on the identification and measurement of inventiveness in mathematics, which form the foundations of this study, are formulated. The primary concern of the study arising from the propositions is stated.

2.2 THE NATURE OF CREATIVITY

Although producers of outstanding work have through the ages been described in various forms reflecting society's appreciation, such as "great" or "creative," considerable difficulties are encountered when attempts are made to define "creativity" operationally. Artists like Michael Angelo, and Picasso, composers like Handel, and Mozart, scientists like Newton and Einstein, poets like Chaucer and Goethe, and others of various disciplines have been enshrined by society as "creative." Yet the scientific study of creativity, its nature, its identification and predictability, developed rather slowly. Guilford draws attention to the initial scarcity and subsequent prolification of publications on creativity:

> In terms of publications on the subject beginning with Galton's Hereditary Genius in 1869 and ending nine decades later, scientific interest has shown an exponential growth. After a painfully slow rate of

> increase in publications, in the late 1930's things began to stir in terms of empirical studies. But it is in the 1950's that the output on the subject virtually exploded. (Guilford, 1967, p. 419).

Various approaches to the identification of creativity may be classified in terms of the identification of product or invention, creative process, and personality characteristics. These approaches complement each other and may be closely interrelated. A (creative) product is generally considered as a sufficient condition for identifying creativity and some writers also consider it a necessary condition. In so far as it is the result of purposeful human energy, it presupposes an inventor who used a creative process to achieve his invention. Not all writers think that a product is necessary for identifying creative talent. While Carl Rogers (1962) maintains that however novel his fantasies may be they must eventuate in a creative product in order to be classified as creative, Frank Barron (1964 p. 112) has made an attempt to identify creativity as a process, independent of a product. For him it is possible to construe creativity as an internal process continually in action but not always observable ... Barron further explains:

> I think of it as something that is happening in the central nervous system. My own basic interest in research on creativity stems from the hope it offers that one may find in psychic creation the same formal variables that can be used to describe creative process in all nature.... (Barron, 1964, p. 113).

Personality characteristics have also been studied as a means of identifying the potential inventor. These have generally been studied in terms of cognitive, motivational and temperamental characteristics of known creative persons, with the expectation that these characteristics would be of use in identifying unknown

creative persons.

2.2-1 The Product

There are several formulations of criteria for identifying creative production, most of them involving "the indispensable kernel of novelty" (Kneller, 1965, p. 3), although Henle (1967) claims that "novelty as such is neither a necessary nor a sufficient condition for creativeness in thinking." (p. 35). Since new productions generally involve previously existing components, the need has been felt for clearly stating the meaning of novelty. Stein (1962) defines novelty in terms of a deviation from previously existing ideas:

> By novel I mean that the creative product did not exist previously in precisely the same form. It stems from reintegration of already existing materials of knowledge, but when it is completed, it contains elements that are new. The novelty of the work depends on the degree to which it deviates from that which exists. (p. 86).

A widely held conception of novelty as a criterion for creativity is that it is sufficient for the product to be new _to the individual_. Kneller thinks that "we create when we discover and express an idea, artifact or form of behavior that is new to _us_" (1965, p. 3). Margaret Mead (1959, p. 223) likewise expresses that "to the point that a person makes, invents, thinks of something that is new to him, he may be said to have performed a creative act." Guilford's measurement of novelty presupposes that "novelty need only apply within the frame of reference of the person himself." This is perhaps what Hadamard (1945, p. 104), is implying when he notes that the student who is attempting to solve a school mathematics problem is engaged on a work of invention:

> Between the work of the student who tries to solve a problem in geometry or algebra and a work of invention, one can say that there is only a difference of degree, a difference of level, both works being of a similar nature.

Kneller also notes a possible difference in degree in the inventiveness of a re-discoverer and an original discoverer. He feels that when a known result is discovered by an individual, the creativity so manifested would be of an inferior order, since the rediscoverer has the advantage denied to the first discoverer, of having grown up in a culture of which the discovery is already a part (Kneller, 1965, p. 3). Kneller's point has considerable merit since for example, the rediscoverer may be aware of tools that considerably facilitate the work of invention, tools which were not at the disposal of the original inventor. If a student who is aware of the methods of the integral calculus in the determination of areas were given the problem of finding a decimal representation of π to any specified degree of precision, he may without much difficulty be able to solve the problem beginning with finding a f(x) such that the area contained within finite limits of x would equal π. He may invent other methods possibly using Taylor series. However this is a problem which originally took centuries to be solved.

The conception of novelty as a criterion for creativity, and as relative to persons, has been used by Guilford (1967, p. 420) to determine a measure for one aspect of originality. His procedure consists of evaluating statistical evidence in respect of an idea that a "reasonably large sample of persons of similar background do not have the idea." The degree of unusualness of the idea in the population of the individual becomes

a measure of originality as measured by novelty.

Various combinations of criteria have been suggested by writers for creative production. Whitting suggests that a creative idea should be new and useful - "an original idea that is also useful, in terms of meeting one of man's needs, is also a creative idea." Henle suggests correctness for "we need some way of distinguishing between the delusions and inventions of the psychotic and the productions of the scientist." She also suggests freedom - which involves freeing oneself from one's ideas in order to solve a problem, - and harmony, which involves reconciliation with the basic structure of the subject matter. (Henle, 1967, p. 32-39). Jackson and Messick (1967), suggest certain 'response properties,' which are unusualness, appropriateness, transformation and condensation. Corresponding to these response properties are 'judgemental standards' - norms, context, constraints, and summary power, and 'aesthetic responses' - surprise, satisfaction, stimulation and savoring. Jackson and Messick feel that while appropriateness and unusualness are considered as necessary criteria for limiting the class of potentially creative products, transformation and condensation are necessary for distinctions of quality and level within the class (Jackson and Messick, 1967, p. 6).

2.2-2 The Creative Process

The problem of what internal processes an inventor uses to invent a product would be considerably simplified if a definite pattern could be shown to exist in the minds of all inventors when a problem is solved creatively. Catherine Patrick feels that there is such a pattern:

> The methods, materials, and aims may vary greatly in the various fields of human endeavour, but the psychological process underlying the production of a work of art, or an invention, or a law, or a scientific formula is fundamentally the same in a-l cases. (Patrick, 1955, p. ix)

Patrick (1955) feels that although we recognize extensive differences between a Shakespearean play, an American folk song, the chemical formula for radium, and a formulated legal cause, the process of thinking which preceded the production of these diverse products exhibited the same essential stages in each instance. Osborn (1957) however warns that:

> those who have studied and practiced creativity realize that its process is necessarily a stop-and-go, catch as catch can-operation -- one which can never be exact enough to rate as science. The most that can be said is that it usually involves some of these phases.

2.2-3 Steps in the Creative Process

Several writers have outlined steps in the creative process. Whitting reports that the steps outlined by Helmholtz are Saturation, Incubation and Illumination. Wallas' (1926) steps are Preparation, Incubation, Illumination and Verification. Osborne's (1957) steps are Orientation, preparation, Analysis, Ideation, Incubation, Synthesis and Evaluation. The steps outlined are generally similar to Wallas's steps, and comment will be made on the steps as formulated by Wallas.

2.2-4 Preparation

During preparation, the thinker acquires as much information as possible about his problem. Helmoltz points out that to bring matters to the point in which "happy ideas come unexpectedly without effort like inspiration" is usually impossible without long preparatory labor." (Patrick, 1955, p. 4). Dewey also stresses that "this bringing

forth of invention, solutions and discoveries rarely occur except to a mind that has previously steeped itself consciously in material relating to its question." Henri Poincare's fifteen days of working on a problem without producing the desired result (Poincare, 1952) may also be considered part of the stage of preparation. He maintains that later inspiration is dependent on such conscious work:

> These sudden inspirations (and the examples already cited sufficiently prove this) never happen except after some days of voluntary effort ... (Poincare, 1952).

The first four stages in Osborn's steps may be said to correspond roughly to the preparation stage. In Orientation, one is setting out, keeping an open mind and being alert to the problem. Preparation involves gathering of pertinent data. Analysis involves breaking down the relevant material and Ideation involves piling up alternatives by way of ideas. Osborne feels that in the ideation stage there should be an uninhibited search for hypotheses, with much emphasis on quantity. He advocates that quantity brings quality.

2.2-5 Incubation

According to Dewey:

> After the mind has ceased to be intent on the problem and consciousness has relaxed its strain, a period of incubation sets in; the material rearranges itself; facts and principles fall into place; what was confused becomes bright and clear; the mixed up becomes orderly, often to such an extent that the problem becomes essentially solved.

This period of incubation has been described by many as a period of unconscious activity. Patrick (1955) points that the stages of preparation and incubation may overlap. Poincare feels that his perparatory effort "set agoing the unconscious machine."

There is some resistance to the acceptance of this stage in

which the unconscious is supposed to operate. Ghiselin (1956) considers that not much is known about this quiescent step and that much of the descriptions reflect "a picturesque substitute for an avowal of ignorance."

2.2-6 Illumination

This is the stage when a fruitful solution comes to mind. It often comes suddenly and generally causes a feeling of certainty and elation. The feeling of certainty is not always justified.

2.2-7 Verification

In this stage, the illumination is verified and revised. As Patrick (1955) puts it, "In this final stage of verification, elaboration or evaluation, the exaggeration of the period of insight are checked against external realities."

Comparatively few attempts have been made to measure the creative process. As Torrance (1962, p. 17) points out, "because of the nature of the creative process and of the limitations of testing situations, only rare attempts have been made to assess the process."

2.2-8 Creativity as Synthesis

It is generally agreed that a mathematical invention is usually a combination of previously existing ideas. Hadamard (1945) maintains that it is obvious that invention or discovery in mathematics or in any other discipline takes place by means of combining ideas. Stein's statement that the creative product stems from a reintegration of previously existing elements is also a support of this view. Einstein (1945) feels that it is a "combinatory play" which seems to be the essence of productive thinking.

While all inventions may be considered as combinations of ideas, not all combinations of ideas are considered as inventions. Hadamard points out that there are numerous possible combinations of ideas, most of which are devoid of interest, a few of which however can be fruitful. Poincare argues that mathematical invention is more than making new combinations, since any one can do that. He maintains that invention consists of choosing the proper combination.

> To create consists precisely in not making useless combinations and in making those which are useful and which are only a small minority. Invention is discernment, choice.

The problem of how the mind makes the appropriate choices is a difficult one. Hadarmard notes that:

> the rules which must guide it "are extremely fine and delicate. It is almost impossible to state them precisely; they are felt rather than formulated...."

Einstein feels that a rather vague combinatory play is involved:

> It is also clear that the desire to arrive finally at logically connected concepts is the emotional basis of this rather vague play with the above mentioned elements. But taken from a psychological viewpoint, this combinatory play seems to be the essential feature in productive thought--- before there is any connection with logical construction in words or other kinds of signs which can be communicated to others. (Einstein, 1945, p. 142).

Wertheimer (1959, p. 238), feels that the mind is energized in some directional way towards the solution of a problem. If S_1 is the state where the thought process starts, and S_2 the situation where the problem is solved, then:

> The thesis is that the very structural features in S_1 with their particular, concrete nature create the vectors, in their direction, quality, intensity, that in turn lead to the steps and operations dynamically in line with the requirements.

Combination of ideas is an integral part of ideational synthesis. Synthesis is generally understood as "unity in a manifold." Bloom's Taxonomy defines it as "the putting together of elements and parts so as to form a whole." Since every combination of ideas involves combining parts, it may be argued that every combination is a synthesis. It is desirable to have some qualification to that synthesis which is not the result of chance, and Wolff's (1963) idea of "Goal Directed Synthesis" seems to be appropriate here. The mind may be conceived as continually engaged in creating combinations of ideas towards a goal, which may be specific or vague, convergent or divergent.

2.2-9 Personality Characteristics

J. P. Guilford, giving a "narrow" definition of creativity in his 1950 inaugural address to the American Psychological Association, said that "in its narrow sense, creativity refers to the abilities most characteristic of creative people." He stressed that the scope of the creativity problem for the psychologist was that of determining the qualities that contribute significantly to a person's producing creative results, which was the problem of creative personality. (Guilford 1962, p. 152). Information in the creative personality has generally been obtained through biographical studies (Cattell, 1959), and in terms of observations on creative persons (Mackinnon, 1967). Various and sometimes conflicting characteristics have been found. Cattell and Bucher (1968) conclude that:

> Neither the biographical nor the empirical studies confirm the "great wits are sure to madness near allied" theory of an association between creativity and neurosis. It appears, on the contrary, that the temperamental stability of eminent scientists in particular is above

> average, though they may often be high in anxiety. For artists, the position is less clear and requires further research.... On the broad second-order factor of introversion-extroversion, it appears that creativity is on the whole more often allied to introversion. But this statement conceals differences between possible patterns of introversion.... The typical personality pattern found in eminent research scientists appears to be one also of high (but not necessarily exceptionally high) intelligence, dominance, desurgent taciturnity, and self-sufficiency. (p. 280).

2.3 GUILFORD'S PSYCHOMETRIC APPROACH

Considerable research has been conducted by Guilford and associates since 1950 into the nature of human abilities, and in particular into creative abilities. The basic approach has been a factor analytic one on creative (cognitive) abilities. While conceding that motivational and temperamental personality traits are of importance in creativity, Guilford and associates believe that the study of creative abilities would aid in determining creative individuals, help in understanding how a creative person thinks as well as give some insight into the creative process. (1965, p. 7).

Guilford defines an individual's personality as his unique pattern of traits. A trait is "any relatively enduring way in which persons differ from one another." The creative personality is thus "a matter of those patterns of traits that are characteristic of creative behavior. A creative pattern is manifest in creative behavior, which includes such activities as inventing, designing, contriving, composing and planning. People who exhibit those types of behavior to a marked degree are recognized as creative." (Guilford, 1962, p. 152-3.)

The Guilford design involved making hypotheses on the existence of creative abilities, constructing and administering tests, and inter-

correlating the scores, to find the underlying abilities that the tests measured. By means of orthogonal factor analytic methods, distinct primary abilities were factored out.

The initial Guilford studies hypothesized at least seven creative abilities--sensitivity to problems, fluency, flexibility, originality, analysis and synthesis, redefinition, and penetration. The results of the analysis were that a factor of sensitivity was found, four factors of fluency -- word, ideational, associational and expressional, two factors of flexibility -- spontaneous and adaptive. No analysis and synthesis factor was found, but redefinition and penetration factors were found. In later factor analytic studies an elaborative thinking factor was found (Guilford, 1965, p. 7).

Guilford's research was in connection with an "Aptitudes Project" at the University of South California. The research was on all intellectual abilities. As a result of his analysis, he classified intellectual abilities in three ways, corresponding to operations, products and contents. Each of these is considered as a parameter, and the component of each parameter is combined to form unique sets of three. Since there are five operations, four kinds of contents, and six kinds of products, there are 120 possibilities, each possibility roughly corresponding to a hypothesised ability.

Divergent production is one of the five operations, and Guilford finds a close relationship between abilities in this category and creative thinking:

> Most of the more obvious contributions to creative thinking are in the divergent production category. The factors of fluency, flexibility, originality and elaboration

> are in that category. It can be said that divergent production abilities are the most direct contributors to creativity. (Guilford, 1965, p. 15).

Guilford and Hoepfner (1967) have identified sixteen divergent production factors in a unified study.

2.4 TESTING FOR INVENTIVENESS IN MATHEMATICS

Carlton (1959) has made an analysis of the educational concepts of fourteen outstanding mathematicians in the areas of mental growth and development, creative thinking, and symbolism and meaning, and she has formulated from their writings, twenty-one characteristics of a potentially creative thinker in mathematics. These characteristics are of considerable value in testing for inventiveness in mathematics. Some of them may profitably be integrated into Guilford's model to identify and measure inventiveness in mathematics. The names at the end of each characteristic are those of the mathematicians whose writings emphasized the characteristic.

<u>Characteristics of the Potentially Creative Thinker (Carlton)</u>

1. An esthetic sensibility, expressed in an appreciation of the harmony, unity, and analogy present in mathematical solutions and proofs and in an appreciation of the structure of the field (Poincare, Hadamard, Gauss).

2. The making up or seeing of problems in data or in situations which arouse no particular curiosity in the other children (Hilbert, Babbage).

3. A desire to improve a proof or the structure of a solution (Boole, Klein, Bocher, Poincare).

4. A seeking for consequences or connections between a problem, proposition, or concept, and what would follow from it (DeMorgan, Whitehead).

5. Desire for working independently of both teacher and other pupils (Gauss, DeMorgan, Whitehead).

6. Pleasure out of communicating concerning mathematics with others of equal ability and interest (Moor; Klein).

7. The speculating or guessing about what would happen if one or more hypotheses of a problem are changed (Poincare).

8. Pleasure derived from adding to the knowledge of the class by producing <u>another</u> solution or <u>another</u> proof beyond those which the class has considered (Whitehead, Miller, Bocher).

9. Pleasure out of working with the symbols of mathematics (Gauss, Whitehead, Poincare, Hadamard).

10. The producing of or conjecturing concerning other meanings for symbols than those the teacher has revealed (Poincare).

11. The making up of mathematical symbols of his own (Babbage, Poincare).

12. The tendency to generalize particular results, either by finding a common thread of induction or by seeing similar patterns by analogy (Poincare).

13. The ability to see a whole solution at one time or to visualize a proof as a whole (Poincare, Hilbert).

14. Intuition as to how things should result (Poincare, Hilbert).

15. A vivid imagination concerning the way things appear in space, the relation of things to each other (Poincare, Hilbert).

16. A vivid imagination concerning the resulting paths or relationships of objects which have motion (Babbage, Poincare).

17. A tendency to speculate concerning unusual applications for the results obtained by the class (DeMorgan, Whitehead).

18. The belief that every problem has a solution (Hilbert).

19. Persistence in working on particularly difficult problems or proofs (Boole, DeMorgan, Miller, Gauss, Babbage, Hadamard).

20. Boredom with repetition or working of a large number of problems dealing with something which he has well in hand (Babbage).

21. Ability to perform many operations without thinking (Whitehead, DeMorgan).

Prouse (1967) constructed a creativity test in mathematics of ten items -- seven in the "divergent thinking category," and three

in the "convergent thinking category." He based his test on Carlton's "characteristics of the potentially creative thinker." Student responses on the divergent thinking problems were scored for fluency and originality. The fluency score was the number of acceptable responses made by the student, and the originality score depended on the frequency of the response in the set of correct responses made on the item (Prouse 1967, p. 876-9). The study tended to indicate a moderate correlation between intelligence test scores and Prouse's creativity test scores, and a high correlation between fluency and originality. Prouse (1967, p. 877), concluded that his study may indicate a need for greater emphasis on divergent thinking approaches in teacher education.

Evans (1964) developed and administered tests to measure the ability to respond in creative mathematical situations at the late elementary and early junior high school level, in terms of fluency, flexibility, and originality. He described fluency as the flow of responses from an individual, and measured it by the number of responses made. He considered flexibility as referring to the variety of responses in a given situation. In scoring flexibility all the responses given by the student were categorized with respect to certain criteria, and one point was given for each category represented in the student's set of responses. Originality was understood as the degree of uncommonness of a given response or kind of response, the originality score for a response being 0, 1, 2, 3, or 4, according to the percentage of examinees who gave the same response. The sum of all the scores for each response represented the originality score for a given test (Evans, 1964, p. 49-51).

Evans reports that all the students who took his tests were able to make responses. He suggests that his testing procedure indicates that "the classroom teacher might provide experiences which enable all of his students at their own level of development, to have part in formulating mathematical concepts." (Evans, 1964, p. 201).

The investigator (Taylor-Pearce, 1969) constructed "divergent thinking" tests in connection with a study of the relative effectiveness of two teaching methods with respect to (divergent thinking) creativity in mathematics at the grade eleven level. The tests were designed to measure fluency, flexibility and originality. The scoring was based on the scoring of Evans (1964). The investigator concluded from the responses of the students that the tests succeeded in encouraging students to formulate mathematical concepts of their own. He called for further research into this type of testing for the teaching/learning situation. He felt that research was needed to reduce or eliminate the subjective aspects of the flexibility score.

Testing for inventiveness or creativity has generally been done by means of problem situations. Polya (1966, p. 126) feels that problems may be classified as "routine" and "nonroutine." "The nonroutine problem demands some degree of creativity and originality from the student, the routine problem does not." Most mathematical inventions have resulted from problems solving activities and the difficult and unsolved problems of mathematics have enabled effective mathematics to be created. Problem making is also an essential part of mathematical inventiveness; problems may be often made in the process of solving mathematical problems. Some problems invented by

Hilbert and Fermat remain unsolved offering rich ground for creative expression. Bell has written that when Klein was asked the secret of mathematical discovery, he replied "You must have a problem. Choose one definite objective and drive towards it. You may never reach your goal, but you will find something of interest in the way."

2.4-1 Testing for Inventiveness

Every production may be considered as a synthesis of ideas, and so far as it is the result of purposeful human activity, it is goal directed. The goal may be convergent or divergent. Divergent production has been advocated as contributing to creativity and the studies reviewed indicate that there is evidence to support the view that divergent production is a promising research approach to the study of inventiveness. The methods of divergent production form the basic methods used in this study. The tests present problem situations which utilize to some extent the principles of Carlton (1959) reported above, and Guilford's "structure-of-intellect" theory.

2.5 FOUNDATIONS OF THE STUDY AND PRIMARY CONCERN

From a study of the literature, the following propositions have been formulated to form the foundations of the study:

(i) That it is possible to evoke and measure inventiveness in mathematics.

(ii) That divergent production in mathematics leads to inventive production in mathematics.

(iii) That individual differences of students in some aspects of inventiveness in mathematics may be ascertained from their individual differences in the components of divergent production.

The primary concern of the study is to demonstrate the validity of the above propositions by validating tests of inventiveness in mathematics constructed by the investigator, using techniques of divergent production. The theoretical foundations of the study are presented in the next chapter, and details are given there of conceptual, psychological, and logical connections between divergent production and creativity, from the writings of Poincare (1952), Osborne (1961), Youltz (1962) and Guilford (1965).

2.6 SUMMARY AND CONCLUSION

A review of the literature on creativity or inventiveness would reveal that there are many unsolved problems concerning the nature of creativity. Many writers consider novelty as "the indispensable kernel" of creativity. The approaches to identifying creativity have been generally through a product or invention, a process, or through personality characteristics. Divergent production has been hypothesized as contributing to creativity and there has been some evidence to support this hypothesis as it relates to mathematics, from the studies of Prouse (1954), Evans (1954) and Taylor-Pearce (1969). The various attempts at solving problems relating to creativity and invention have led to the formulation of theoretical views which have highlighted and illumined aspects of the subject.

Creativity is of great importance to learning and culture. It may also be vital to confidence and self-image. Guilford's divergent production approaches have been selected as promising and adaptable to the school situation. The methods for evoking and measuring inventiveness used in this study are based on these approaches.

CHAPTER III

THEORETICAL FOUNDATIONS AND THE FORMULATION OF HYPOTHESES

3.1 INTRODUCTION

The purpose of this study was to establish principles for constructing and scoring tests of inventiveness in mathematics at the senior high school level. The study was centered around the validation of tests of inventiveness in mathematics constructed by the investigator for students undertaking a particular mathematics course (Math 20) in Edmonton high schools. The significance of the problem has been discussed in the first chapter of this dissertation. In the second chapter, a review has been made of the literature concerned with the identification and measurement of inventiveness. In this chapter, theoretical principles underlying the construction and validation of the tests are discussed, and the hypotheses of the study are presented.

3.1-1 Letter Symbols

A number of letter symbols are used in this and subsequent chapters to denote expressions used frequently in this study. An alphabetical key to these symbols may be found in Table 1 of appendix A.

3.2 DIVERGENT PRODUCTION

The tests used in this study to evoke inventiveness, are tests of divergent-production (DP), in the sense that Guilford (1967) uses the term. He explains that "divergent production is a concept

defined in accordance with a set of factors of intellectual ability that pertain primarily to information retrieval and their tests, which call for a number of varied responses to each new item." (Guilford, 1967, p. 138). He notes that DP tests are conspicuously absent from modern group tests of intelligence particularly after machine scoring came into existence. Guilford points out an important distinguishing feature of DP tests: "divergent production tests require the examinees to produce their own answers, not to choose from alternatives given to them." (Guilford, 1967, p. 138).

Guilford distinguishes between divergent and convergent production: "Convergent production rather than divergent production is the prevailing function when the input information is sufficient to determine a unique answer." He makes a deliberate contrast between the two types of production in which extreme points of difference are emphasised. This contrast is set side by side in tabular form below. The words used are Guilford's. The form is the investigator's

<u>Divergent Production</u>	<u>Convergent Production</u>
1.1. The problem itself may be loose and broad in its requirements for solutions.	1. The problem may be rigorously structured and is so structured, and an answer is forthcoming without much hesitation.
1.2. If the problem is properly structured, the individual may have an incomplete grasp of it.	
1.3 The individual may have a complete grasp of the problem, but unable to find the unique answer immediately, resorting to trial and error behavior, which means divergent production alternated with evaluation.	
2. Restrictions are few	2. Restrictions are many.

Divergent Production	Convergent Production
3. The search is broad.	3. The search is narrow.
4. Output is in quantity.	4. Output is limited.
5. Criteria for success are vague and somewhat lax and may, indeed stress variety and quantity.	5. Criteria are sharper, more rigorous, and demanding.

Guilford recognizes that a middle course between these contrasted types of functions is quite common in every day life. "The individual very frequently engages in much divergent production on the way to a convergent answer, as he puzzles over a mathematical problem and he tries one solution after another." (Guilford, 1967, p. 215).

3.2-1 Divergent Production and Creativity

Guilford finds a close connection between divergent production and creativity: "It can be said that divergent production abilities are the most direct contributors to creativity" (Guilford, 1965, p. 15). Similar opinions may be implied from the writings of Poincare (1952), Osborn (1962), and Youltz (1962). Poincare states:

> In fact, what is mathematical creation? .... To create consists precisely in not making useless combinations and in making those which are useful and which are only a small minority. Invention is discernment, choice ....

Although Poincare stresses that many combinations of ideas may be useless, he is stating that there is a useful minority of combinations which could be considered as invention. Amplifying his contention that invention is choice he continues:

> To invent, I have said is to choose, but the word is perhaps not wholly exact .... The sterile combinations do not even present themselves to the mind of the inventor. Never in the field of

> his consciousness do combinations appear that are not really useful, except some that he rejects but which have to some extent the characteristics of useful combinations.

Poincare may be interpreted as affirming that useful divergent productions are formed in the mind of the inventor, giving him the opportunity of selecting the most appropriate for his purpose. Poincare also highlights variety when he notes that "Among chosen combinations, the most fertile will often be those formed of elements drawn from domains which are far apart." (Poincare, 1962, p. 35, 36).

Alex F. Osborn of whom it has been said (Parnes and Harding, 1962, p. 19) "if there is one person who has contributed most to the development of the creative problem solving movement over the longest span of years, it is Alex F. Osborn," states that the production of ideas leads to the production of ideas of quality.

> In ideative effort, quantity breeds quality. Until recently we could substantiate this principle only on the basis of the laws of probability plus empirical evidence. The principle is now confirmed by scientific research which found that those who thought up twice as many ideas thought up <u>more</u> than twice as many good ideas in the same length of time.

Youltz (1962, p. 195, 196) feels that when a person has exhausted habitual solutions to a problem he is then faced with the necessity of making a new solution. He may be hampered by his habits here, and he may or may not be able to break out of them:

> He tries to do something new, but finds that each thing he does is a familiar pattern. When he has done familiar things repeatedly with no success, he may with great effort take parts of different habitual acts and combine these parts into new action.

3.2-2 Divergent Production in "Structure-of-Intellect" Theory

Guilford has developed a unified theory of intelligence, and the model he has developed in this connection is generally referred to as the "Structure of Intellect" model. He has classified intellectual abilities in three ways or dimensions, corresponding to operations, products and contents. There are five operations, four kinds of contents, and six kinds of products. These are given below, as are the corresponding descriptive symbols as used by Guilford (1967b, p. 426).

OPERATIONS	CONTENTS	PRODUCTS
Cognition (C)	Figural (F)	Units (U)
Memory (M)	Symbolic (S)	Classes (C)
Divergent Production (D)	Semantic (M)	Relations (R)
Convergent Production (N)	Behavioral (B)	Systems (S)
Evaluation (E)		Transformations (T)
		Implication (I)

Combining each of the five operations, with each of the four kinds of content and each of the six kinds of products gives 120 triples. Each triple is basically considered as a hypothesized ability. It may be readily seen that corresponding to divergent production there should be twenty-four triples. Guilford and associates (Guilford and Hoepfner, 1966) have identified sixteen DP factors.

3.3 ADAPTATIONS OF GUILFORD'S THEORY FOR THE PURPOSE OF THIS STUDY

This study is concerned with evoking and measuring subject-related inventiveness. The subject matter that the students have learned in the high school is an essential aspect of the study. It is the mathematics that the students have acquired that they are expected to use in an inventive manner. No distinction is made

between the various components of content that Guilford distinguishes, in regard to the kinds of tests constructed for the study. The contents of the students' responses are expected to be mathematical, involving some or all of these components. Product distinctions are made. The various product categories are used as goals for the construction of tests. Descriptions of the abilities considered to be involved in these product categories have been adapted from Guilford and Hoepner (1966) and are given below:

ABILITY		DESCRIPTION
1. Divergent Production of Mathematical Units	(DMaU)	The ability to produce various elementary mathematical ideas related to a mathematical situation.
2. Divergent Production of Mathematical Classes	(DMaC)	The ability to resist fixedness in mathematical thinking and to produce mathematical ideas that are different in relation to a mathematical situation.
3. Divergent Production of Mathematical Relations	(DMaR)	The ability to produce or recognize mathematical relationships.
4. Divergent Production of Mathematical Systems	(DMaS)	The ability to organize elementary mathematical ideas into complex ones.
5. Divergent Production of Mathematical Transformations	(DMaT)	The ability to produce original responses involving re-interpretations and re-definitions.
6. Divergent Production of Mathematical Implications	(DMaI)	The ability to produce mathematical implications from a given set of conditions.

3.4 VALIDATION

The central aspect of this study is the validation of tests of inventiveness in mathematics constructed by the investigator. The

validation procedures adopted are to some extent based on the "Standards for Educational Tests and Manuals," which have been approved by the governing bodies of the American Psychological Association (APA), the American Educational Research Association (AERA), and the National Council on Measurement in Education (NCME) (Jackson and Messick, 1967, p. 169-189). This document will be referred to in this dissertation as the <u>Standards</u>.

The supreme importance of validity in test construction has been stressed in recent years. Cattell and Bucher (1968, p. 92) point this out strongly:

> Traditionally, in examining any test, one begins by asking how reliable it is and then proceeds to ask if it is also valid. But if a test has no validity, and no utility ..., one need not waste time studying its reliability. A still more compelling reason for studying validity first ... is that reliability is kept in its proper perspective if we consider it primarily as a modifier of validity. Similarly, in buying a mechanical instrument, we may wish to know whether it is lasting and dependable, whether it will rust, for example, or fall to pieces on a second use; yet one's primary concern is whether the tool will do the job in mind.

Cronbach (1970) states a similar view: "The quality that most affects a test is its validity ... No matter how satisfactory it is in other respects, a test that measures the wrong things is worthless." Validity has often been conceived in terms of the extent to which a test measures what it is supposed to measure (Guilford, 1954, Anastasi, 1961). Gulliksen (1950) defines validity as "the correlation of the test with some criterion." He states that a test could in this sense have many validities, and that validity cannot be regarded as a fixed and unitary characteristic of a test. The position stated by Cattell and Bucher (1968, p. 92) is that "validity

in a generic sense is the ability of a test to predict some behavioral measure other than itself." The <u>Standards</u> explain validity in terms of achieving certain aims: "Validity information indicates the degree to which the test is capable of achieving certain aims."

Considerable dissatisfaction has been expressed as to the sufficiency of the "correlation with a criterion" definition of validity. Thurstone's statement in this regard is telling: "In the field of intelligence tests, it used to be common to define validity as the correlation between a test score and some outside criterion. We have reached a stage of sophistication where the test-criterion correlation is too coarse. It is obsolete." Loevinger (1967, p. 79) contends that this type of validity "is not a suitable basic concept for test theory: it does not provide an adequate basis for test construction."

The <u>Standards</u> distinguish three "of the rather numerous aims of testing":

1. The test user wishes to determine how an individual performs at present in a universe of situations that the test situation is claimed to represent.

2. The test user wishes to forecast an individual's future standing or estimate an individual's present standing on some variable of particular significance that is different from the test.

3. The test user wishes to infer the degree to which the individual possesses some hypothetical trait or quality (construct) presumed to be reflected in the test performance.

In accordance with the three aims, the <u>Standards</u> name three aspects of validity corresponding to these aims: content validity, criterion-related validity, and construct validity. Content validity is demonstrated by showing how well the content of the test samples the

class situations or subject matter about which conclusions are to be drawn; criterion-related validity is demonstrated by comparing the test scores with one or more external variables considered to provide a measure of the characteristics of the behavior in question, and construct validity is ordinarily studied when the tester wishes to increase his understanding of the psychological qualities being measured by the tests.

Considerable discussion has been centered around the concept of construct validity since it appeared in the 1954 APA Technical Recommendations. Basic descriptions have been given by Cronbach and Meehl (1955). Constructs should be set within "nomological networks" or "interlocking system of laws which constitute a theory," and construct validation "is only possible when some of the statements in the network lead to predicted relations among observables." (Cronbach and Meehl, 1955, p. 300). While Loevinger (1967) supports the concept of construct validity: "construct validity is the whole of validity from a theoretical point of view," she warns against a confusion of constructs and traits. She makes a distinction between construct and traits analogous to the distinction between statistic and parameter: "The trait is what we aim to understand and the corresponding construct represents our current best estimate of it." Campbell (1960) recommends that two types of validity be distinguished - trait validity and nomological validity. Some writers like Bechtoldt (1959) have criticised the concept of construct validity adversely and Michel (1964, p. 59) warns about the danger of a misuse of the concept of construct validity: "Unverkennbar is aber die Gefahr, dass der Begriff „Konstructvalidität"

missbrauchlich verwendet wird, um das Fehlen sauberer Validitatsuntersuchungen zu kashieren." ("There is undoubtedly a danger that the notion of "construct validity" is misused to fill in for the absense of a purer validity research.")

The Standards explain that the three aspects of validity are only conceptually independent, and that information on each apsect should normally be involved in a complete study (APA, 1967, p. 178). Attempts have been made to utilize all three aspects (content, construct, criterion) of validity in this study, and the hypothesis of the study reflect this.

Another aspect of great importance in test validation is reliability. This is a measure of consistency, as Maguire and Hazlett (1969, p. 118) have pointed out: "The fundamental concept of reliability is consistency not correlation." Catell and Bucher (1968, p. 98), point out that "consistency is the most important property of a test after validity, from which it is conceptually different," but feel that the concept of consistency is in some ways wider than reliability (p. 98). There are various domains in which consistency may appropriately be studied. These domains include raters, time, culture, personality characteristics, socio-economic conditions, and environment.

Both validity and reliability are essential basic requisites for satisfactory tests. A test without consistency across a domain of interest is of little use in situations involving that domain even if it could be shown that the test is valid. Validity without consistency is also of little use. The consistency studies in this work are generally of the homogeneity type, indicating the consistency

with which various raters agree in determining a common construct.

3.5 Formulation of Hypotheses

3.5-1 Purpose of the Hypotheses

Hypotheses were used in this study to facilitate the validation investigations. They were formulated to represent aspects of the process of validation and to reflect expectations such that decisions on the validity of the tests could be made from the experimental verification or refutation of these expectations.

3.5-2 Content Validation

The procedure adopted in content validation was to present a group of specialists in mathematics, mathematics education, and measurement, with a set of constructed tests of certain defined abilities, and to request them independently to rate the tests on their suitability to test the ability as defined. Their ratings were expressed in symbols which denoted whether the tests were "very good," "good," "satisfactory" or "unsatisfactory" tests of the abilities defined. A basic condition for accepting a test as a suitable test for an ability was that the average rating of the specialists as judges should indicate that it was a "satisfactory" test of the ability. A more powerful situation was considered to be the situation in which every specialist as judge considered the test as at least a "satisfactory" test of the ability.

The above considerations led to the following hypotheses:

Hypothesis 1a That each test of an ability presented to the judges will receive an average rating from the judges, indicating that it is at least a "satisfactory" test of the ability.

Hypothesis 1b That each test of an ability presented to the judges

will receive an assessment from each judge indicating that it is at least a "satisfactory" test of the ability.

3.5-3 Construct Validation

The procedure adopted in construct validation was to investigate empirically a set of expectations of critical importance that would logically result if certain assumptions were valid. These assumptions were based on the principles underlying the constructing the tests, and the anticipated structure of the tests.

The tests of divergent production were constructed to measure abilities associated with product categories of Guilford's "structure-of-intellect" model. Three conceptually distinct measures were obtained for most tests. These were measures of facility in production, variety in production, and novelty in production, and these are referred to in this dissertation as "DP measures."

It was expected that the tests which had satisfied the requirements for content validation had been "suitably classified." If they had been suitably classified, they would be expected to measure the same constructs. Using the principles of factor analysis (Harman, 1967), this would mean that the variables of such tests should determine the same common factor. In this case, a rotation to simple structure of the test battery containing the "suitably classified" test variables, while reducing the complexity of the variables of the battery, should result in a simple factor structure in which "suitably classified" tests determine at least one common factor. The above considerations were the bases of hypotheses 2a, 2b and 2c.

Hypothesis 2a That test measures of facility in production will reveal a simple factor structure such that tests hypothetically classified in the same product category determine the same factor.

Hypothesis 2b That test measures of variety in production will reveal a simple factor structure such that tests hypothetically classified in the same product category determine the same factor.

Hypothesis 2c That test measures of novelty in production will reveal a simple factor structure such that tests hypothetically classified in the same product category determine the same factor.

One of the basic principles used in constructing the DP tests was that they should be subject-related. A logical expectation from this would be that the subject-related DP tests in mathematics should significantly predict school achievement in mathematics, where "achievement" was indicative of subject mastery. This consideration was the basis of hypothesis 3.

Hypothesis 3 That divergent production abilities in school mathematics predict school achievement in mathematics significantly.

A distinction is made in this dissertation between divergent problem solving and convergent problem solving. All the DP tests constructed by the investigator for this study are examples of divergent problem solving tests. Since convergent problem solving tests in mathematics are considered by many as measures of inventiveness in mathematics, the relationship between divergent problem solving and convergent problem solving was investigated in the following hypothesis.

Hypothesis 4. That divergent production abilities in mathematics significantly predict convergent problem solving ability in mathematics.

Three DP measures were obtained from the same test performance on each test, except for "classes" test from which only two measures were obtained. These were measures of facility, variety, and novelty in production. Facility in production is the ability to produce a quantity of responses, variety in production is the ability to produce different responses spontaneously, and novelty in production is the ability to produce rare responses. Tests in the classes category were marked for variety and novelty only.

Osborne's (1962) idea that "quantity breeds quality," and Poincare's (1952) affirmation that "among the most fertile combinations will often be those formed from domains which are apart," lead to the expectation of an inter-relatedness among the DP measures, beyond that which would normally be expected from the fact that they are "experimentally dependent" (Thurstone, 1947, p. 441). A hierarchical relationship is hypothesized among these measures:

Hypothesis 5. That DP measures of production may be hierarchically ordered within each product category.

Guilford (1967, p. 63) states that "the order along each dimension of the model has some logical reasons behind it but without any great deal of compulsion." He regards units as basic, and although he lists implications last, he thinks that transformations may have a valid claim for that position. He feels that the position of implications may be best next to units:

> There might be some sense in putting implications immediately below units, since implications are the simplest and most general way in which units can be connected.

> There is reason for putting systems below units and relations, since both enter into systems; but implications do also.

An investigation of critical importance was to determine the hierarchical relationships among the product categories, and in particular to determine whether Guilford's hypothesized "logical" order was empirically plausible. The above considerations were the bases of hypothesis 6.

Hypothesis 6. That product categories may be hierarchically ordered within each DP measure.

3.5-4 Criterion Validation

The criterion for inventiveness used in this study was the mean assessment of the inventiveness of a response production by a group of experts. This group was called the SIGNIFICANT GROUP the name being taken from Stein's (1962, p. 86) definition of a creative work as "a novel work which is accepted as tenable or useful or satisfying by a significant group of others at some point in time." The position taken in this study is that if a production in a field of study is considered by a significant group of experts in the field of study at some point of time, as indicative of inventiveness in the producer, then it is an inventive production.

All the hypotheses formulated in connection with criterion validation made use of the criterion for inventiveness of the study in some ways, but not always in the "correlation with a criterion" sense. Some of the hypotheses may with some justification have been classified with construct validation.

One of the propositions which formed the foundations of the study (Section 2.5) was that divergent production in mathematics would

lead to inventiveness in mathematics. The validity of this proposition was investigated in hypotheses 7 and 8.

Hypothesis 7. That each subject will produce at least one inventive response.

Hypothesis 8. That each test will evoke inventive responses.

Another of the fundamental propositions of the study was that individual differences of students in some aspects of inventiveness may be ascertained from their differences in the components of divergent production. The validity of this proposition is investigated in hypothesis 9.

Hypothesis 9. That each DP measure correlates significantly with the criterion measure of inventiveness.

The novelty score for a response was calculated in terms of the statistical rarity of the response. Since novelty is considered by many as the "indispensable kernel" of creativity, it would be expected that there would be significant agreement between measures of novelty and inventiveness of a response production. The above considerations were the bases of hypothesis 10.

Hypothesis 10. That the novelty measures for responses and the inventiveness measures for responses correlate significantly.

3.6 SUMMARY AND CONCLUSION

Divergent production tests in mathematics are developed in this study to evoke inventive responses from high school mathematics students. Theoretical considerations lead to the expectation that these methods will afford an effective way of evoking and measuring inventiveness in mathematics.

The validation of the tests take the form of investigations into their content, construct, and criterion validities. Several hypotheses are formulated which reflect aspects of the process of validation. The criterion for inventiveness used is the pooled assessment of a significant group of experts of the response productions of students.

CHAPTER IV

TESTING AND MEASURING

4.1 INTRODUCTION

This study was undertaken with the purpose of establishing principles for constructing and scoring tests of inventiveness in mathematics in the senior high school. The primary concern of the study was to validate tests of inventiveness in mathematics constructed by the investigator for students taking a senior high school mathematics course (Math 20) in Edmonton. Theoretical principles underlying the construction and validation of the tests, and the hypotheses of the study were discussed in chapter III. In this chapter, the final form of the tests, and the experimental design of the study will be presented.

4.2 CONSTRUCTION AND SELECTION

4.2-1 Construction and Content Validation

The main principles used in constructing the DP tests were as follows:

1. Each test should present a mathematical problem situation and request more than one response production to it.

2. Several correct answers should be conceivably possible.

3. Response productions of high quality should be conceivably possible

4. The tests should be such that in an open book examination, each student should be capable of making at least one correct response.

5. The subject matter of each test should be based on the Math 20 course. In addition, the student should also have opportunity to use mathematics that he knows, whether he learned it in school or otherwise.

6. Each test should be aimed at testing one and only one of the abilities DMaU, DMaC, DMaR, DMaS, DMaT, DMaI, as defined in section 3.3.

Tests of convergent problem solving (CPS) were adapted by the investigator from three problems. One was a problem from a Russian Olympiad, and the other two were problems found in the Hungarian Problem Solving Book (Rapaport, 1963).

The DP and CPS tests constructed by the investigator were evaluated on their face and content validities in testing defined abilities by six judges. Each judge was a university professor, possessing a doctorate degree. Of the six judges, one was a specialist in measurement, three were specialists in mathematics education, and two specialized in mathematics. The tests selected for the preliminary study were based on the evaluation of the judges. Details of the evaluation may be found in the analytical investigations connected with hypotheses 1(a) and 1(b) in chapter VI

4.2-2 Preliminary Study and Final Form of Tests

A preliminary study was conducted at an Edmonton high school in January, 1970 and the final form of the tests was selected as a result of the study. A report of the preliminary study and the rationale for the final selection of tests are presented in Appendix B.

The DP tests which were used in the main studies were as follows:

Test No.	Test	Product Category
I	Write down as many mathematically true statements as you can about a <u>Shola</u> in the sense defined below: A <u>Shola</u> is an odd integer divisible by 39	Units
1.	Write down as many mathematically true statements as you can about an <u>Epudom</u> in the sense defined below: An <u>Epudom</u> is an integer divisible by 35	Units
IB	Write down as many mathematical statements as you can about the following function. Each statement should be such that it is true or would be true under certain conditions. Try to make each statement represent one main idea only. Use your imagination. $y = 2^x, x \in R.$	Units
1B	Write down as many mathematical statements as you can about the following function. Each statement should be such that it is true or would be true under certain conditions. Try to make each statement represent one main idea only. Use your imagination. $y = 2^x(x^2-5x+6), x \in R.$	Units
II	Invent as many systems of equation as you can such that the solution set of each includes the number (1,2,3). Try to make the systems as different from each other as possible. When you have thought out a pattern for making sequences, give two or three examples of the pattern, and group the similar systems together. Then look for a different pattern and group in a similar way. Please indicate the groups of systems that are different.	Classes
2	Invent as many systems of equations as you can such that the solution set for each system is (5, 7). Try to make the systems as different from each other as possible. When you have thought out a pattern for making sequences, give two or three examples of the pattern, and group the similar systems together. Then look for a different	Classes

Test No.	Test	Product Category
2(cont'd)	pattern and group in a similar way. Please indicate the groups of systems that are different.	
IIB	Think out and write down different sets of integers (m, n, q) such that $m^2+n^2 = q^4$. Try to make the sets of triples as varied as possible. When you have found a pattern of triples, give two or three examples of the pattern, and group the similar triples together. Then look for a different pattern and group in a similar way. Please indicate the groups of triples that are different.	Classes
IV	Show that if b and c are real, and x_1, x_2 are the roots of the quadratic equation $x^2+bx+c = 0$, then $x_1+x_2 = -b$, using as many distinct methods as you can. When you have thought out a method, write down an outline of the method, such that it would be possible to see how you would proceed if you had time. Then, look for another method. Try to think out and write outlines of as many methods as you can.	Systems
4.	Show that $\frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2}$, using as many distinct methods as you can.	Systems
4B	Invent several sequences of numbers such that in each case the generating pattern is as complex and unusual as you can make it.	Systems
5.	Imagine that you wish to explain to Grade Six students why we cannot divide by zero. Think out some unusual mathematical approaches that you can use. List as many of these approaches as you can.	Transformations
5B	Invent different operations on numbers each of which behaves in an unusual way. In each case define the operation carefully. Try to invent as many operations as you can.	Transformations
VI	Suppose that you are working in a system in which it is true that 2 ⊗ 8 = 4. Think out mathematical statements that would be true in this system, and in	Implications

Test No.	Test	Product Category
VI (cont'd)	each case explain briefly (as far as you can) why.	
6B	T(n) is defined as the set of all positive integers less than n which divide n evenly. Think out and guess as many properties of T(n) as you can.	Implications

The CPS tests which were used in the main studies were as follows:

Test No.	Test
VII	Find all two digit numbers, x, such that the product of the digits equals $x^2 - 10x - 10$.
VIIB	Find all positive numbers x, such that $x(x+3)$ is the square of an integer.
7B	Find all two digit natural numbers, y, such that the sum of the digits of y is $y^2 - 10y - 9$.

4.2-3 Administrative considerations

The preliminary study influenced the following decisions concerning administration in the main studies.

1. The students should be informed of the purposes of the test, the theory underlying the investigation on inventiveness, and be given such information as would encourage them to participate fully in the investigation. They should be given opportunity for discussion and questioning prior to the administration of the tests.

2. That the investigator should be the sole administrator of the tests.

4.3 MAIN STUDIES

Two main studies were conducted in two Edmonton high schools. Both studies were administered in April, 1970. The first main study will be referred to as Main Study 1, and the second as Main Study 2.

4.3-1 Sample for Main Study 1

The sample for main study 1 consisted of forty volunteers in an Edmonton high school, who took all the tests of the study. The investigator invited Math 20 students of this school to take

part in the experiment during non-teaching periods, and fifty-five volunteers took at least one of the tests, but only forty took all the tests. This sample of forty volunteers will be referred to as main sample 1.

4.3-2 Sample for Main Study 2

The sample for main study 2 consisted of 62 math 20 students in an Edmonton high school who took all the tests during class hours. These were regular students in four Grade Eleven classes with a total enrolment of 91.

4.3-3 Administration of Main Study 1

The tests were administered by the investigator in a classroom at the school mainly during the noon hour, and the twelve tests administered were completed by most of the students in six consecutive days, two being administered each day. A few students took the tests after school. Students who missed one day were allowed to continue the tests on another day. The investigator devoted the day before the first test to discussing the purpose of the tests, and matters generally concerned with the experiment, giving the students opportunity for discussion and questioning. Ten minutes were allowed for each DP test, and fifteen minutes were allowed for each CPS test. The twelve tests administered to the sample were composed of ten DP tests and two CPS tests. The ten DP tests were designed to measure abilities in five product categories -- units, classes, systems, transformations, and implications. There were two tests for each category. The administered tests were part of the tests listed in section 4.2-2. Tests 1 and I were given as unit tests, tests 2 and II as classes tests,

tests 4 and IV as systems test, tests 5 and 5B as transformations tests, tests VI and 6B as implications tests, and tests VII and 7B as CPS tests.

The students appeared motivated and most of them appeared keen to take the tests. The forty volunteers were willing to use some of their non-teaching hours for an extended period to take the tests, and for most of them this was done in twenty-minute periods during their noon hour. The students worked hard at the tests, and as far as was apparent to the investigator, they applied themselves consciencíously to the tests.

4.3-4 Administration of Main Study 2

The tests were administered by the investigator in the four classes during class hours. Twelve tests were administered to the sample. Ten of the twelve tests were DP tests and the remaining two were CPS tests. The ten DP tests were expected to measure abilities in three product categories -- units, classes, and systems. There were four tests in the units category, tests 1, I, 1B, and IB, three in the classes category, tests 2, II, and IIB, and three in the systems category, tests IV, 4, and 4B. The CPS tests administered to main sample 2 were tests VII and VIIB.

The first meeting with the students was considered an orientation meeting. The investigator explained the purpose of the tests, the theory underlying the investigation on inventiveness, and answered questions from the students. Some time was also spent in reviewing the Math 20 subject matter relevant to the tests. The testing was done within three additional periods, each period lasting fifty-eight minutes. The students appeared keen to do their best,

and applied themselves diligently to the tests.

4.3-5 Limitations on Product Categories Investigated

The product categories investigated in main study 1 were limited to units, classes, systems, transformations, and implications, and did not include relations. Those investigated in main study 2 were units, classes and systems. Ten DP tests were administered in each study, and six of the ten DP tests of main sample 2 were the same as the units, classes and systems tests of main study 1.

4.4 SCORES

4.4-1 Ratings from the SIGNIFICANT GROUP

The SIGNIFICANT GROUP was a group of experts whose opinions on the inventiveness of the response productions were considered as the main criterion for inventiveness. The name SIGNIFICANT GROUP was taken from Stein's (1962, p. 86) definition of a creative work as a novel work which is acceptable as tenable or useful or satisfying to a significant group of others at some point in time (The emphases are the writer's). The SIGNIFICANT GROUP consisted of three university professors in mathematics, four university professors in mathematics education, one university professor in measurement, and one principal of a junior high school. All the professors had doctorate degrees, and all were actively engaged in research. The principal had an M.Ed. degree. Each member of the SIGNIFICANT GROUP was asked to rate standardized student response productions in accordance with his answer to the following question:

Is the response indicative of inventiveness in the high school student producing it?

The letter to each judge then requested:

If your answer to the question is No, kindly rate the response .. 0

If your answer is Yes, kindly rate as follows:

The response is indicative of a low degree of inventiveness .. 1

The response is indicative of a low/high degree of inventiveness .. 2

The response is indicative of a high degree of inventiveness .. 3

The response is indicative of a very high degree of inventiveness 4

4.4-2 Scores from DP Tests

Three different types of scores were obtained from each test, with the exception of DMaC tests. These were facility scores, variety scores and novelty scores. Only variety and novelty scores were obtained for the DMaC tests.

4.4-2-1 Facility Scores

Facility scores correspond to fluency scores in the investigator's earlier study (Taylor-Pearce, 1969, p. 52). One facility mark was awarded for each appropriate response. An appropriate response was a response which satisfied the requirements of a problem. An appropriate response may be thought of as a "correct" response, one which fits the qualifications of a question. Minor arithmetic mistakes do not make a response inappropriate. Each response listed in Appendix C is an appropriate response to the test for which it is listed.

4.4-2-2 Variety Scores

Variety scores correspond generally to Guilford's spontaneous flexibility scores (Guilford and Hoepner, 1966). A student's variety score was computed as the number of different appropriate responses he produces to a problem situation in a unit of time.

The number of different responses was determined in accordance with a technique developed by the investigator, and presented in the next chapter.

4.4-2-3 Novelty Scores

A novelty score was assigned to each response, depending on the degree of uncommonness of the response in the totality of responses made by students in the sample. The principles for determining the novelty score of a response were developed from principles of Guilford and associates (Wilson, Guilford, and Christenson, 1962). The developments are reported in the next chapter.

4.4-3 Scores from Convergent Problem Solving (CPS)Tests

Polya (1957) has distinguished four phases involved in problem solving: Understanding the problem, devising a plan, carrying out the plan, and looking back. The investigator devised a marking scheme for the CPS tests which was based on Polya's phases. The test scorer investigates the following questions with regard to a student's presentation on a problem.

UNDERSTANDING THE PROBLEM

Did the student indicate expressly or implicitly that he had at least a partial understanding of the problem? Yes/No

Did the student indicate expressly or implicitly that he had a complete understanding of the problem? Yes/No

DESIGN

Did the student give evidence expressly or implicitly that he had a design to solve the problem? Yes/No

Was the design such as would possibly lead to a complete solution?
Yes/No

PROCEDURE

Did the student show some mathematical competence in the pursuit of his design? Yes/No

Did he discover significant relationships which could effectively lead to a solution of the problem? Yes/No

Did he effectively use these relationships to obtain a solution? Yes/No

SOLUTION

Did the student obtain a partial solution? Yes/No

Did he indicate that a complete solution exists? Yes/No

Did he obtain a mathematically complete solution? Yes/No

Various weights may be given to the affirmative and negative answers. In this study, equal weights were given to each Yes, each being rated as one, and each No being rated as zero. The student's score was rated as the sum of the ten component scores.

4.4-4 Scores indicating subject matter mastery

The scores indicating subject matter mastery in Main sample 1 was obtained from a multiple choice achievement test set by the school to all Math 20 students. The Kuder-Richardson 20 reliability coefficient of this test was 0.76. The test had 30 items and was taken by a total of 254 Math 20 students. The scores indicating subject matter mastery in Main sample 2 was obtained from an achievement test constructed by the investigator and accepted by the school

as an achievement test for the school. The test had 32 items and was taken by a total of 92 Math 20 students. The Kuder-Richardson 20 reliability coefficient was .70.

4.4-5 I.Q. Scores

I.Q. Scores I.Q. scores were obtained from the respective schools. In main sample 1, the I.Q. scores obtained were the Lorge-Thorndike verbal and non-verbal scores. Most of the scores were obtained in 1968, with some in 1967 and 1969. The scores in main sample 2 were obtained on the California Test of Mental Maturity in 1966.

4.5 MARKING AND ANALYZING

The tests were marked and revised by the investigator as examiner. Three university students assisted the investigator by re-marking samples of the tests.

The test scores were analyzed by the investigator with the aid of the APL/360 computing facilities of the University of Alberta.

4.6 SUMMARY AND CONCLUSION

Tests of divergent production in high school mathematics were constructed by the investigator to evoke inventiveness in senior high school mathematics students. They were tried out in a preliminary study, and a final set was determined. The final was composed of tests which were designed to test abilities in terms of some of the product categories in Guilford's structure-of-intellect theory.

A sample of the tests were administered to student volunteers in an Edmonton high school, and another sample was administered to regular students in another high school in Edmonton. An orientation meeting was held in each administration of the tests. The tests were

in each case administered by the investigator as sole administrator.

Scores from the tests were obtained in terms of facility in production, variety in production, and novelty in production. Other scores were obtained which were relevant to the validation of the tests.

All tests were marked by the investigator, and the test scores were analyzed by the investigator with the aid of the APL/360 computing facilities of the University of Alberta.

CHAPTER V

MEASURING SPONTANEOUS FLEXIBILITY AND ORIGINALITY

5.1 INTRODUCTION

The study was undertaken with the purpose of establishing principles for constructing and scoring tests of inventiveness in senior high school mathematics. The primary concern of this study was to validate tests of divergent production in mathematics which were constructed to evoke and lead to the measurement of inventiveness in mathematics. Three measurements of divergent production were defined in the previous chapter. The facility measure was designed to measure fluency, the variety measure was designed to measure spontaneous flexibility, and the novelty measure was designed to measure originality. In this chapter, theoretical developments leading to techniques for the determination of the variety and novelty measures of test of divergent production are presented.

5.2 THEORETICAL CONSIDERATIONS

Problems attendant on the objectivity of the spontaneous flexibility measure have been raised by May and Metcalf (1965) and the investigator (Taylor-Pearce, 1969, p. 117). Guilford defines the ability involved in spontaneous flexibility as "the ability to produce a variety of class ideas appropriate to a given idea," and explains that the adjective "spontaneous" is used because "the thinker is flexible even when he has no need to be," and this

is contrasted with "adaptive flexibility" in which the thinker would fail to solve the problem if he were not flexible (Guilford, 1962, p. 158). The problem of determining a score for spontaneous flexibility is a problem of classifying maximally similar responses into distinct classes or "clusters." Determining which statements of ideas are different and hence belong to distinct classes, often presents great difficulties. Statements are clearly either the same in all respects or different in some respect. Statements may express similar mathematical ideas in semantically dissimilar ways, or even in mathematically dissimilar ways. Two mathematical statements, P. and Q., may be equivalent in the sense of "P if and only if Q," and yet the implication may be profound, and hence someone who thinks out P may not readily or even ultimately think out Q. In this case P and Q may logically belong to different classes of ideas although they are intrinsically equivalent.

Another difficulty arises from the fact that the same set of ideas may be classified in several ways, each possibly yielding its own number of different ideas, depending on the classification principle adopted. It is also possible that a particular classification principle may never lead to an invariant number of distinct classes, when applied to the same set of ideas, and this may lead to difficulties in determining the number of different ideas. Thus variety scores obtained using different principles of classification may not be compatible, and there is need to show that an adopted procedure for determining different ideas leads to at least a conceptually invariant number.

Classification principles may be developed for each particular test, as was done in the investigator's earlier study (Taylor-Pearce, 1969), but this procedure may not lead to results which are comparable between tests. It would be more desirable to have classification techniques which are applicable to all DP mathematics tests. Secondly, it would be desirable to have principles which are logically valid in the sense that they are based on some theoretical principles, and represent a plausible way of differentiating between ideas. Thirdly, one would wish that the number of different ideas obtained by these techniques could if possible be proved to be invariant, so that deviation among test scorers could logically be attributed to error.

5.2-1 Theoretical Foundations of the Technique

The technique developed is the outworking of developments on adaptations of theories due to De Bono (1969) and Guttman (1954). The technique was developed by the investigator in the process of marking several divergent production responses.

De Bono distinguishes between two types of thinking, vertical thinking, and lateral thinking:

> You can dig a hole in a different place by digging the same hole deeper. Vertical thinking is concerned with digging the same hole deeper. Lateral thinking is concerned with digging the hole somewhere else. The aim of both is effectiveness.

De Bono feels that vertical thinking has been generally encouraged by education, and that the mind uses vertical thinking naturally: "not only does the mind use vertical thinking naturally (albeit inefficiently) but it is trained to use it by education." Yet, according to De Bono, while vertical thinking is concerned with the

development and use of ideas, it is lateral thinking which is concerned with making new ideas (De Bono, 1969, p. 160).

De Bono contrasts lateral thinking with vertical thinking: Vertical thinking "is essentially sequential in nature. One proceeds step by step along a path," but lateral thinking "does not have to be sequential." Vertical thinking "is based on the principle that one must not be wrong ... Yet the fear of being wrong is the biggest bar to new ideas." Vertical thinking "chooses the most promising approach, singles it out and follows it as far as it goes," but lateral thinking "is not interested in single approaches, no matter how promising they may be." Outside influences are excluded in vertical thinking, but in lateral thinking "one realizes that the disruption of a particular fixed idea may only come through a random intrusion so one not only welcomes such intrusions but actively seeks to generate them." Vertical thinking "tends to build up large established patterns since the use of large patterns speeds up both communication and information processing," but lateral thinking "seeks to break down established patterns into small units. One seeks to disrupt patterns so that the information released may re-form itself into new and better patterns."

De Bono's descriptions of vertical and lateral thinking are highly suggestive of rigid and flexible thinking. The essence of the factor of spontaneous flexibility according to Guilford (1967, p. 325) is "the readiness to shift from class to class." Flexible thinking implies a shift, a jump, an unconnectedness in ideas while rigid thinking implies a sequence, a pathway, a connectedness in ideas. It is not only in flexbile thinking that different

ideas are produced, for as De Bono notes "You can dig a hole in a different place by digging the hole deeper." However, where the purpose is to measure flexibility, this type of difference is "non-relevant." If we assume that a student's productions in a DP test may be used to identify the types of thinking that he employs in the test situation, then two non-vertical responses would be considered as having "relevant" differences, with the underlying assumption that all non-vertical responses are flexible. Classifying the responses according to "non-relevant" differences would possibly lead to a measure of the "relevant" classes in which the responses fall.

In outlining his radex theory, Guttman (1954) identifies and assigns order to two notions of differences between tests which could be applied to differences in vertical-type thinking and lateral-type thinking. He distinguishes between "kind" and "degree of complexity." Limiting his development to the "simplest case of the radex theory which can be completely portrayed by a simple two-dimensional diagram," he deduces that "within all tests of the same kind, differences will be in degree." Also, "all tests of the same degree of complexity will differ among themselves only in the kind of ability they define."

Guttman's notion of complexity has been adapted by the writer to provide a method of identifying vertical-type or path-type ideas. He explains complexity as follows:

> Suppose we are given n tests, $t_1, t_2, \ldots, t_n$, which differ on a single complexity factor ... Test t_1 is the least complex. Test t_2 is next; it requires everything t_1 does and more .... In general test t_{j+1} is more complex than t_j, and hence

> requires what all the preceeding tests require, plus something more.

Guttman's complexity distinctions among tests are adapted to complexity distinctions among mathematical concepts. If given a mathematical concept A a concept-set A is defined as the set of all mathematical properties necessary to establish A, then it may be said that A is more complex than B if concept-set B is a subset of concept-set A. Using set-theoretic notation, this may be written as $B \subset A$, or equivalently, $B - A = \emptyset$.

5.3 DEVELOPMENT OF VARIETY SCORES

5.3-1 Criterion Concept of a DP response production

The technique was developed by the investigator in the process of marking the DP responses of the study. Each student response to a test situation was expected to be a simple statement. When a compound statement was used, it was treated as a composite of simple statements provided that this treatment did not distort the meaning of the statement. The criterion for differentiating the responses of the study was the <u>subject matter domains</u> of the responses. This choice was influenced by Poincaré's suggestion: "Among the combinations, the most fertile will often be formed from domains which are apart." The "concept of the subject-matter-domain" of a response (statement) was used as the <u>criterion mathematical concept</u> for determining different responses. This concept was conceived as that relationship which would cause a set of entities to be identified in that domain. The following notation was used in this development for a student's response statement, and the criterion mathematical concept of a response.

<u>Notation</u> Let Q_i denote a student's response statement. (Natural number i is used to identify the particular response).

<u>Notation</u> Let $C(Q_i)$ denote the criterion mathematical concept of response Q_i.

5.3-2 <u>Theoretical Establishment of the Existence and Uniqueness of a variety score for Spontaneous Flexibility, given the prior establishment of the criterion mathematical concepts of response productions</u>

The idea of the concept-set of a concept is introduced in the following definition:

<u>Definition</u> Given a mathematical concept $C(Q_i)$, the concept set of $C(Q_i)$ is defined to be the set of mathematical properties which are necessary to establish the concept $C(Q_i)$.

<u>Notation</u> Let $\underline{Q}_i$ denote the concept-set of $C(Q_i)$.

A property p is an element of $\underline{Q}_i$ if it is necessary to establish $C(Q_i)$, as $\underline{Q}_i$ is defined above. This may be expressed in another way. If whenever a set of entities possesses a relationship expressed by a concept, $C(Q_i)$, it necessarily possesses the relationship expressed by the property p, then p is an element of $\underline{Q}_i$. Thus the property of groups is an element of the concept set of fields, and the property of closure is an element of the concept set of groups. The property of divisibility is an element of the concept set of divisibility, the property of linearity is an element of the concept set of polygons, the transitive property is an element of the concept set of equivalent relations, but the property of ordered fields is not an element of the concept set of complex numbers.

Set theoretic notation is used in the development below. Attention is called in particular to the following true statements:

1. For any sets X and Y, $X - Y = \emptyset \Longleftrightarrow X \subset Y$.
2. If X and Y are distinct, then $X - Y = \emptyset \Longrightarrow Y - X \neq \emptyset$.

<u>Path connection and Independence -- Definitions</u>

Two concept sets $\underline{Q}_1$ and $\underline{Q}_2$ are considered as <u>path-connected</u> if one of them is a subset of the other. They are considered as <u>independent</u> if $\underline{Q}_1 \cap \underline{Q}_2 = \emptyset$. Two mathematical concepts are considered as <u>path-connected</u> or <u>independent</u> if their concept-sets are path-connected or independent. Two responses Q_1 and Q_2 are considered as <u>path-connected</u> or <u>independent</u> according as $\underline{Q}_1$ and $\underline{Q}_2$ are path-connected or independent.

<u>Maximally Reduced Subuniverse</u>

A possible line of investigation in the quest for a unique score for spontaneous flexibility is suggested by the problem of obtaining minimal bases for mathematical structures. Here, one would seek -- possibly by a method of successive reduction, the smallest possible set of responses that represent the total generalized information of all the responses, such that no two members of this smallest set are path-connected. The quest for such a smallest set leads to the determination of a maximally reduced subuniverse, which is defined as follows: Let U be a universe containing n distinct sets, $\underline{Q}_1, \underline{Q}_2, \ldots, \underline{Q}_n$. Let U_r be a subuniverse of U. Then U_r is called a <u>maximally reduced subuniverse</u> of U if

1. Every element in U is a subset of at least one element in U_r.
2. No element in U_r is a subset of any other element in U_r.

The existence and uniqueness of a maximally reduced subuniverse for any finite universe U are established in Theorems 1 and 3.

Theorem 1 There is at most one maximally reduced subuniverse for any given universe.

Proof Let U_t and U_r be two maximally reduced subuniverses of a given universe U. Let $\underline{Q}_{t_i}$ be any element in U_t. Then $\underline{Q}_{t_i}$ being also an element in U is a subset of at least one set $\underline{Q}_{r_j}$ in U_r. Since $\underline{Q}_{r_j}$ has to be a subset of some element $\underline{Q}_{t_m}$ in U_t, it follows that $\underline{Q}_{t_i} \subset \underline{Q}_{r_j} \subset \underline{Q}_{t_m}$ and hence $\underline{Q}_{t_i}$ is a subset of $\underline{Q}_{t_m}$. Since $\underline{Q}_{t_i}$ cannot be a subset of any other element in U_t, it follows that $\underline{Q}_{t_i}$ and $\underline{Q}_{t_m}$ are the same. Hence $\underline{Q}_{t_i} = \underline{Q}_{r_j} = \underline{Q}_{t_m}$. Hence $U_t \subset U_r$. Similarly, it can be shown that $U_r \subset U_t$. Hence $U_t = U_r$.

A Representative of an element in a Universe

The notion of a representative $R(\underline{Q}_i)$ of an element $\underline{Q}_i$ in a universe U is useful in this development. It is defined as follows: For any element $\underline{Q}_i$ in a universe U, a representative of the element $\underline{Q}_i$ is an element $R(\underline{Q}_i)$ in U, such that $\underline{Q}_i$ is a subset of $R(\underline{Q}_i)$, and $R(\underline{Q}_i)$ is a subset of no other element in U. The next theorem establishes that every element in a universe containing a finite number or sets, has at least one representative.

Theorem 2. Every element in a universe containing n distinct sets has at least one representative.

Proof Let U be a universe containing n distinct sets. Let $\underline{Q}_{k_1}$ be an arbitrary element in U. If it is possible to select an element

Q_{k_2} in U - $\{Q_{k_1}\}$ such that Q_{k_1} is a subset of Q_{k_2}, the process of selection may be continued. Otherwise it is stopped. If it is possible to select an element Q_{k_3} in U - $\{Q_{k_1}, Q_{k_2}\}$ such that Q_{k_2} is a subset of Q_{k_3}, the process may be continued. Otherwise it is stopped. The process is to progressively select a Q_{k_i}, in the subuniverse U - $\{Q_{k_1}, Q_{k_2}, Q_{k_3}, \ldots, Q_{k_{i-1}}\}$, where i progressively takes values from 2 to n, such that $Q_{k_{i-1}} \subset Q_{k_i}$, as long as it is possible to continue the process of selection. The number of elements in U - $\{Q_{k_1}, Q_{k_2}, Q_{k_{i-1}}\}$ becomes smaller by one each time the process continues and thus there can be at most n-1 selections after Q_{k_1} has been arbitrarily chosen. Hence there is a last Q_{k_m}, such that no element may be found in
U - $\{Q_{k_1}, Q_{k_2}, Q_{k_3}, \ldots, Q_{k_m}\}$,
such that Q_{k_m} is a subset of that element. Since

$Q_{k_1} \subset Q_{k_2} \subset Q_{k_3} \subset \cdots \subset Q_{k_m}$, it follows that Q_{k_1} is a subset of Q_{k_m}, and since Q_{k_m} contains the sets $Q_{k_1}, Q_{k_2}, \ldots Q_{k_{m-1}}$, it follows that Q_{k_m} is a subset of no other element in U. Hence Q_{k_m} is a representative of Q_{k_1}.

The next theorem establishes that there is at least one maximally reduced subuniverse for any universe containing n distinct sets.

Theorem 3 There is at least one maximally reduced subuniverse of any universe containing n distinct sets.

Proof Let $U = \{Q_1, Q_2, Q_3, Q_4, \ldots Q_n\}$.

Let $L = \{Q_i \in U: Q_i - Q_j = \emptyset\}$ for some $Q_j \in U$, $i, j, = 1, 2, 3, \ldots, n$, $i \neq j$,

It is to be proved that (U-L) is a maximally reduced subuniverse.

This will be done by showing that:

1. (U-L) is not empty.

2. Every element in U is a subset of at least one element in (U-L).

3. No element in (U-L) is a subset of any other element in (U-L).

Proof Let Q_k be any element in U. Then Q_k has at least one representative in U. Let Q_{k_r} be a representative of Q_k. Then since Q_{k_r} is a subset of no other element in U, it cannot be an element in L and hence is an element in (U-L). Hence (U-L) is not empty, and since Q_k is a subset of Q_{k_r}, it follows that every element in U is a subset of at least one element in (U-L). The third aspect of the proof follows immediately. If two distinct elements of (U-L) are such that one is a subset of the other, this would imply that one of them is an element of L, which is a contradiction of the assumption that they are elements in (U-L). Hence (U-L) is a maximally reduced subuniverse.

It follows from theorems 1 and 3 that for any given universe containing n distinct sets, there is one and only one maximally reduced subuniverse. The corollary affords another way of looking at the maximally reduced subuniverse.

Corollary. The set of all representatives of all elements in a universe containing n distinct sets, is the maximally reduced subuniverse of

the universe.

Proof Let R be the set of all representatives of all elements in a universe U containing n sets. Then from the proof of theorem 3, it is clear that R is contained in (U-L). Any element of (U-L) has to be its own representative since otherwise it would be an element of L. Hence R = (U-L).

5.3-3 Variety Score

It follows from theorems 1 and 3 that there is one and only one maximally reduced subuniverse for a given universe containing n distinct sets. Hence for every universe of n concept sets, there exists an invariant number $r \leq n$, such that r is the number of distinct elements in the maximally reduced subuniverse. For a particular student his set of n_1 distinct responses to a DP test may be used to subjectively determine a set of n distinct criterion-mathematical-concept-sets. This set constitutes a universe of sets, and an invariant number r exists for the number of elements in the maximally reduced subuniverse. This number r may be taken as the variety score.

It is to be noted, however, that although no two of the concepts in the maximally reduced subuniverse are path-connected, they are not necessarily independent. If it is considered essential to obtain independent concepts as evidence of flexibility, then it may not be possible to keep the concepts as intact representations of the responses. The following procedure may be adopted leading also to a conceptually invariant number: Let $U_r = \underset{-}{Q}_1, \underset{-}{Q}_2, \ldots, \underset{-}{Q}_r$, be a maximally reduced subuniverse of a universe U, containing n sets. Then, the specific part of $\underset{-}{Q}_i$ is defined to be:

$$S(\underline{Q}_i) = \underline{Q}_i - (\underline{Q}_1 \cup \underline{Q}_2 \cup \cdots \cup \underline{Q}_{i-1} \cup \underline{Q}_{i+1} \cup \cdots \cup \underline{Q}_r)$$

The specific part of the response Q_i are those properties of Q_i, which are not repeated in any generalized concept. The set of all specific parts is a universe containing disjoint sets. This may be characterized as follows: $S = \left\{ S(\underline{Q}_i) \ : \ \underline{Q}_i \in U \right\}$. It should be pointed out that in general, S is not a subuniverse of U. It should also be noted that whenever a $\underline{Q}_i$ is contained in the totality of the remaining r-1 concept sets, the specific part of $\underline{Q}_i$ is the empty set. Let the number of distinct elements in S be denoted by r_1. Then $r_1 \leq r \leq n$.

<u>Representative-concepts and Specific-concepts Variety Scores</u>

The r score above will be referred to as the <u>representative-concepts variety score</u>, and the r_1 score as the <u>specific-concepts variety score</u>.

5.4 TECHNIQUE

One way of obtaining the representative-concepts and the specific-concepts variety scores is to proceed as follows:

<u>Stage</u>	<u>Procedure</u>
1.	Obtain the concept set $\underline{Q}_i$ for each statement Q_i.
2.	Ensure that only distinct concept sets are retained.
3.	Successively discard any concept set that is contained in another, doing this by logical analyses of mathematical relationships. Obtain a <u>possible maximally reduced subuniverse</u>.
4.	Verify whether the obtained possible maximally reduced subuniverse is the maximally reduced subuniverse using a two dimensional matrix table, in which the ith row and

jth column of the matrix of results contains the answer to the question "Is the ith concept set a subset of the jth concept set?". An affirmative result is indicated by a 1, and a negative by a 0. The maximally reduced subuniverse is obtained if, and only if, the resulting r x r matrix contains ones in the diagonal, and zeros everywhere else. The representative-concepts variety score is then found to be r.

5. The specific-concepts variety score is obtained by obtaining the number of concept sets in the maximally reduced subuniverse whose specific parts are non-empty.

5.4-1 Limitations of the Technique

A limitation of the technique arises from the fact that the determination of the criterion mathematical concept of a response depends on the judgement of the test scorer. The recommended criterion mathematical concept here is the concept of the subject-matter-domain of the response. This refers to the generalized subject matter areas from which the mathematical content of the response was presumably drawn. It seems reasonable to expect substantial agreement among test scorers on what the subject-matter-domain of a response is.

5.4-2 Illustration of the technique in action

The technique is illustrated below with the following selected responses to test 1.

Test 1. Write down as many mathematically true statements as you can about an Epudom in the sense defined below:

An Epudom is an integer divisible by 35.

Selected Student Responses

Q_1: Every Epudom is divisible by 5.

Q_2: Every Epudom is divisible by 7.

Q_3: Epudoms are not primes.

Q_4: The set of Epudoms is infinite.

Q_5: Epudoms will always end in 5 or 0.

Q_6: The sum of the first and last digits of an Epudom will never exceed 14.

Q_7: Epudoms are closed with respect to addition and multiplication.

Q_8: The set of Epudoms form a number system.

Q_9: Epudoms do not form a field.

Q_{10}: There are as many Epudoms on one side of zero as on the other.

Q_{11}: Epudoms can have an infinite number of digits.

The above statements are a sample of the student responses listed in Appendix C. Variety scores for the above set of responses will now be determined in the stages of Section 5.4.

Stage 1.

The concept set of each statement is determined in the first stage. The subject matter domain of each statement is first determined. In determining the subject matter domain of a statement, every attempt is made to generalize the subject matter, without attributing the domain to a more complex domain than is absolutely necessary. Numbers and particular examples should not be included in the conception of the domain.

The subject matter domains of the eleven statements may be determined as follows:

Statement	Subject-Matter-Domain
Q_1	Divisibility
Q_2	Divisibility
Q_3	Divisibility
Q_4	Infinity
Q_5	Decimal Representation and Divisibility
Q_6	Decimal Representation, divisibility and Boundedness
Q_7	Closure
Q_8	Number Systems
Q_9	Fields
Q_{10}	One-one correspondence
Q_{11}	Infinity, and Decimal Representation.

The concept sets ($\underline{Q}_i$'s) may accordingly be determined as follows:

$\underline{Q}_1$: The concept set of <u>divisibility</u>.

$\underline{Q}_2$: The concept set of <u>divisibility</u>.

$\underline{Q}_3$: The concept set of <u>divisibility</u>.

$\underline{Q}_4$: The concept set of <u>infinity</u>.

$\underline{Q}_5$: The concept set of <u>decimal representation</u> and <u>divisibility</u>

$\underline{Q}_6$: The concept set of <u>decimal representation</u>, <u>divisibility</u> and <u>boundedness</u>.

$\underline{Q}_7$: The concept set of <u>closure</u>.

$\underline{Q}_8$: The concept set of <u>number systems</u>.

$\underline{Q}_9$: The concept set of <u>fields</u>.

$\underline{Q}_{10}$: The concept set of <u>one-one correspondence</u>.

$\underline{Q}_{11}$: The concept set of <u>infinity</u> and <u>decmial representation</u>.

<u>Stage 2</u>.

The test scorer ensures that only distinct concept sets are retained. Since $\underline{Q}_1 = \underline{Q}_2 = \underline{Q}_3$, and no other concept sets are judged to be equal, there are now nine distinct concept sets:

$\underline{Q}_1$, $\underline{Q}_4$, $\underline{Q}_5$, $\underline{Q}_6$, $\underline{Q}_7$, $\underline{Q}_8$, $\underline{Q}_9$, $\underline{Q}_{10}$, and $\underline{Q}_{11}$.

<u>Stage 3</u>.

The test scorer now discards any concept set that is contained in another, doing this by logical analyses of mathematical relationships.

From a consideration of the concepts, it becomes evident that (1) $\underline{Q}_1 \subset \underline{Q}_5$,

(2) $\underline{Q}_4 \subset \underline{Q}_{11}$,

(3) $\underline{Q}_5 \subset \underline{Q}_6$,

(4) $\underline{Q}_7 \subset \underline{Q}_8$,

(5) $\underline{Q}_7 \subset \underline{Q}_9$,

(6) $\underline{Q}_8 \subset \underline{Q}_9$,

(7) $\underline{Q}_5 \subset \underline{Q}_{11}$, and

(8) $\underline{Q}_1 \subset \underline{Q}_6$.

The concept sets $\underline{Q}_1$, $\underline{Q}_4$, $\underline{Q}_5$, $\underline{Q}_7$ and $\underline{Q}_8$ are therefore discarded, and the concept sets $\underline{Q}_6$, $\underline{Q}_9$, $\underline{Q}_{10}$, and $\underline{Q}_{11}$ are retained as forming a <u>possible maximally reduced subuniverse</u>.

<u>Stage 4</u>.

During this stage, the <u>possible maximally reduced subuniverse</u>

is examined in tabular form to verify whether it is a maximally reduced subuniverse of representatives. (See corollary to theorem 3). Each entry in the ith row and jth column of the table corresponds to the answer to the question "Is the ith concept set a subset of the jth concept set?" An affirmative reply is indicated by a 1, and a negative reply by a 0.

$\underline{Q}_i \subset \underline{Q}_j$?	$\underline{Q}_6$	$\underline{Q}_9$	$\underline{Q}_{10}$	$\underline{Q}_{11}$
$\underline{Q}_6$	1	0	0	0
$\underline{Q}_9$	0	1	0	0
$\underline{Q}_{10}$	0	0	1	0
$\underline{Q}_{11}$	0	0	0	1

The representative concepts variety score (r) is 4.

Stage 5.

At this stage, the specific concepts variety score may be obtained. The idea of the specific part of the concept set $\underline{Q}_i$, $S(\underline{Q}_i)$, (see section 5.3-3) is used here. The problem here is to determine whether the concept set $\underline{Q}_i$ is contained in the union of the other remaining concept sets or not. For the purposes of obtaining the specific concepts variety score, it is sufficient to determine whether the specific part of $\underline{Q}_i$ is, or is not empty. The problem is investigated: "Is $S(\underline{Q}_i) \neq \emptyset$?" The number of affirmative answers is the specific-concepts variety score.

A logical analysis of the maximally reduced subuniverse to determine specific-concepts score for the maximally reduced subuniverse derived above, results in the following:

Statement	$S(\underline{Q}_i) \neq \emptyset$
$\underline{Q}_6$	1
$\underline{Q}_9$	1
$\underline{Q}_{10}$	1
$\underline{Q}_{11}$	1

It follows that the specific-concepts variety score is 4.

5.5 VARIETY PROCEDURE

A subject's variety score in a test was an estimate of the number of distinct representatives of his responses where each representative was determined in terms of all the responses made by all subjects in that test situation.

5.6 NOVELTY SCORES

5.6-1 Novelty Score for each Response

A novelty score was assigned to each response, depending on the degree of uncommonness of the response in the totality of responses made by the students in the sample. The score for each response was determined according to the following procedure: Let s represent the novelty score for a response Q_i. Then an upper bound for s is set in accordance with the proportion of the number of students who made responses classified as being the same as Q_i, as follows:

Proportion of number of students making response to total number of students in the sample	Score
.81 - 1.00	0
.61 - .80	$s \leq 1$
.41 - .60	$s \leq 2$
.21 - .40	$s \leq 3$
.00 - .20	$s \leq 4$

A response Q_i is awarded its upper bound score if, and only if, there is no response Q_j, having a lower upper-bound, such that

the concept set $\underline{Q}_i$ is a proper subset of the concept set $\underline{Q}_j$. Otherwise, the upper bound of response Q_i is lowered to the upper-bound of response Q_j. The procedure is continued until each response can be awarded their upper bound score.

5.6-2 Rationale for the Procedure

The use of frequency proportions to determine novelty is based on Guilford's (1967b, p. 420) rationale for operationally determining novelty:

> Novelty need apply only within the frame of reference of the person himself. If we say that an idea is novel only if no one before has ever achieved it, we are completely blocked; for we could never hope to establish the fact, one way or the other. We have some possibility of establishing whether or not the idea is novel for the individual by knowing his past history. We rarely have enough information regarding his past history, however, to be certain. Operationally, indirect evidence can be found by demonstrating that a reasonably large sample of persons of similar background do not have the idea. In other words, we establish the fact that his idea is unique in his population, where his population can be sampled in a reasonable number. (Guilford, 1969b, p. 420).

The procedure of partitioning the frequency proportion into fifths, adopted here appears to be just one convenient way of doing this, and was adopted from Evans (1964.) Wilson, Guilford and Christensen (1962) describe a method which uses a similar partitioning.

One aspect of this procedure, however, that does not <u>appear</u> to have been accounted for by previous investigators, is that it is not only the most novel responses that may have statistically low frequencies in a sample. Some responses may appear so insignificant to a student, that he does not bother to write them down, but he may write down responses which indicate that he knows these ideas and more. When most students ignore the insignificant, but appropriate responses, these responses tend to have low frequencies.

It was considered that the frequency proportion could only determine upper bounds for the responses, and that if there is evidence that a response is implicit in another response, the response should not be assigned a novelty score higher than the novelty score of the response in which it is subsumed. These considerations led to the above technique.

5.6-3 Novelty Score of a Student in a DP test.

The novelty score of a student in a test was the mean of the two highest novelty scores that he obtained on his responses. It was considered that a measure of the best production of a student in a test was what was required to determine how novel he could be. The highest response score could be the best measure of this, but the next highest was also used in order to reduce error in measurement.

5.7 SUMMARY AND CONCLUSION

It has been pointed out in this chapter that there are serious difficulties attendant on the determination of a score for the number of different responses made by a student in a test of divergent production in mathematics. A technique has been developed for determining this (variety) score, and the technique is based on certain theoretical considerations of different types of thinking.

Two variety scores have been developed, a representative concepts variety score, and a specific concepts variety score. It has been shown that under certain conditions, these scores exist, and are invariant.

The consideration that some trivial as well as highly inventive responses may be rare in a population of responses has led to

the adoption of a procedure which is an amendment on the novelty procedure of Guilford.

CHAPTER VI

TEST VALIDATION: ANALYTICAL INVESTIGATIONS

6.1 INTRODUCTION AND OVERVIEW

This study was undertaken with the purpose of establishing principles for constructing and scoring tests of inventiveness in mathematics at the senior high school. The primary concern of the study was to validate tests of inventiveness in mathematics constructed by the investigator for students taking a high school mathematics course (Math 20) in Edmonton. The hypotheses of the study were presented in chapter III. In this chapter, analytical processes and procedures leading to decisions on the hypotheses of the study are reported.

6.1-2 Nature of Report

6.1-2-1 Order of Investigation. Analytical investigations concerning content, construct, and criterion validities were conducted, and are reported here. The order in which the investigations were carried out was to a large extent necessitated by a desire for a systematic development leading to the determination of the largest subset of the tests which met the most critical expectations for validation. Thus the tests which were investigated in connection with construct validation were those tests which had satisfied the requirements for content validation. In the construct validation studies, it was the tests which had been shown to be "suitably classified" on the basis of the factor analytic studies that were used in the

investigations of relationships between divergent production abilities and abilities involved in subject mastery and convergent problem solving. Further, it was the tests which had shown stability over the DP measures on the basis of the factor analytic studies, that were used in hierarchical investigations.

6.1-2-2 Analytical Procedures. Conventional analytical procedures used in this study are reported briefly, and appropriate references are supplied. It was found necessary to adopt certain analytical procedures which were to some extent developed by the investigator. These developments are reported in detail in this chapter in connection with the investigations which necessitated them. The best example of this concerns the hierarchical investigations conducted in connection with construct validation. It is often possible to obtain a trivial hierarchy for any set of test variables at will. This can be done by utilizing the procedures of diagonal factorization (Harman, 1967, p. 102), and pivoting on appropriate variables. Accordingly, where interest is in a non-trivial hierarchy, it is desirable to specify the type of hierarchy expected, so that if it exists, it could exist in only one way. The hierarchical analytical development reported here was directed toward this goal. The procedures involve principles due to Guttman (1954) and Burt (1954), and also utilize certain mathematical principles and procedures, which are reported by Moise (1963). These procedures result in a development which would reveal a betweeness-complexity hierarchy in two possible arrangements, such that the order of the variables in one arrangement is the opposite of the order of the variables in the other arrangement.

6.1-2-3 Presentation. The process of decision making on the hypotheses is presented in three stages. These are (1) Preliminary Discussion, (ii) Analysis and Results, and (iii) Discussion.

In the first stage of preliminary discussion, background information is presented on the hypothesis or hypotheses being considered. Theoretical considerations underlying the analysis may be presented at this stage.

In the second stage, statistical procedures relevant to decision making on the hypotheses are presented, and a summary of the analysis and results is given.

In the third stage, the results are discussed, and further decisions may be made.

6.1-3 Letter Symbols

A number of letter symbols are used in this and other chapters to denote expressions frequently used in this study. An alphabetical key to these symbols may be found in Table 1 of Appendix A.

6.2 CONTENT VALIDATION

The hypotheses tested in connection with content validation were:

Hypothesis 1a That each test of an ability presented to the judges will receive an average rating from the judges, indicating that it is considered at least a "satisfactory" test of the ability.

Hypothesis 1b That each test of an ability presented to the judges will receive an assessment from each judge indicating that it is at least a "satisfactory" test of the ability.

6.2-1 Preliminary Discussion

Forty-six tests of seven defined abilities were developed by the investigator and rated by six judges on their face and content validities. Details of information given to the judges, and the original forty-six tests may be found in Appendix E. The judges assigned symbols V, G, S, and U to each test, according as it was considered a "very good," "good," "satisfactory," or "unsatisfactory" test of the ability that it was expected to measure.

6.2-2 Analysis and Results

A summary of the ratings of the judges on the appropriateness of each of the forty-six tests is given in table 1 of Appendix E. A summary of the analysis of the ratings is given in Table I

Each of hypotheses 1a and 1b was treated as a hypothesis for each of the forty-six tests. In testing for hypothesis 1a, natural numbers 3, 2, 1, and 0 were assigned to the ratings V, G, S, and U, respectively, reflecting their underlying order relationship. The average rating (AR) was the mean of the ratings of the six judges, found by multiplying the numbers of V, G, S, and U ratings by 3, 2, 1, and 0 respectively, and dividing the result by six.

The decision rule for testing hypothesis 1a was to reject the hypothesis for a test if AR for that test was less than one, and not to reject it for that test otherwise. As may be seen from table 1 of Appendix E, and also from table I, the hypothesis was rejected for only one of the forty-six tests. Forty-five out of forty-six tests satisfied the basic condition for content validation.

TABLE I

SUMMARY OF ANALYSIS OF JUDGES' RATINGS ON APPROPRIATENESS OF TESTS[1]

Ability	N(J)	N(S)	N(P)
DMaU	10	10	9
DMaC	9	9	4
DMaR	6	6	4
DMaS	2	1	0
DMaT	6	6	1
DMaI	6	6	2
CPS	7	7	7
Total	46	45	27

[1]N(J) denotes the number of tests judged.

N(S) denotes the number of tests with at least a "satisfactory" average rating.

N(P) denotes the number of tests which were considered by every judge as at least "satisfactory."

The decision rule for testing hypothesis 1b was to reject the hypothesis for a test if any one of the judges rated that test as "unsatisfactory," and not to reject the hypothesis for that test otherwise. As may be deduced from table 1 of Appendix D, and also from table I, the hypothesis was rejected for nineteen out of forty-six tests. Twenty seven out of forty-six tests were assessed as appropriate by each judge for the ability they were expected to measure.

6.2-3 Discussion

Test selection for the preliminary study was made on the basis of the content validation study. It was considered that a test with a higher average rating should be selected in preference to a test with lower average rating, subject to the condition that no test with an average of 1 or less should be selected, and subject also to the condition that wherever possible, tests should be selected which best appeared to measure the ability they defined.

No DMaS test was selected. On the advise of some of the judges, tests in other categories were adapted to serve as tests of systems. All other tests selected satisfied the initial basic condition for content validation. A record of the tests selected for the preliminary study may be found in table 2 of Appendix D.

6.3 CONSTRUCT VALIDATION

Hypotheses 2a, 2b, and 2c were tested in order to provide analytical evidence for evaluating the appropriateness of the classification of the tests into product categories.

Hypothesis 2a That test measures of facility in production will

reveal a simple factor structure in which tests hypothetically classified in the same product category determine the same factor.

Hypothesis 2b That test measures of variety in production will reveal a simple factor structure in which tests hypothetically classified in the same product category determine the same factor.

Hypothesis 2c That test measures of novelty in production will reveal a simple factor structure in which tests hypothetically classified in the same product category determine the same factor.

6.3-1 Preliminary Discussion

The hypotheses were tested to investigate the appropriateness of the classification of the tests in the same product categories. Tests which primarily measure the same ability were expected to determine at least one common factor in a factor analysis of the correlation matrix of test variables, when a simple structure solution was obtained.

The simple structure solution used in this study was an othorgonal simple structure solution from an initial principal axis solution, using Kaiser's varimax criterion. Harman (1967, p. 294) advocates that "This procedure not only does a better job of approximating the classical simple structure principles, but it also tends to lead to factorially invariant solutions."

Since the differences in the abilities which the tests measured could be hierarchical as well as independent, it was not essential that tests in different product categories should determine different factors.

Since the three DP measures were "experimentally dependent" in the sense that they were all obtained from "the same performance,"

they were not factor analyzed in the same test battery. Thurstone (1947, p. 442) advised that "it is best to avoid inserting in a test battery two or more measures that are taken from the same test performance."

Following Guttman's (1956) development and assessment of a "best possible systematic estimate of communality," squared multiple correlations (SMC's) were used as communalities. The number of common factors was accordingly determined as the least number of factors of the reduced correlation matrix with SMC's as communality estimates that accounted for at least the trace of the matrix. Since a principal axis factoring was carried out this was readily obtainable by determining the smallest number of latent roots which accounted for the sum of the SMC's.

6.3-2 Analysis and Results - Facility Measures of Main Sample 1

Facility measures for main sample 1 were obtained in eight DP tests, two in each of the product categories, DMaU, DMaS, DMaT, and DMaI. Tests in the classes product category were designed to measure variety and novelty, but not facility. The analysis also included two CPS (convergent problem solving) test measures, Lorge-Thorndike verbal (LTV) and non-verbal (LTNV) intelligence test measures obtained from the school, and an achievement test measure obtained from the school and designated here as a subject mastery test (SMM1). The latter tests were included to provide reference for the interpretation of factors. A varimax solution was obtained from an initial principal axis solution of the matrix of intercorrelation of the thirteen test variables with squared multiple correlations in the diagonal. The thirteen latent roots obtained were: 5.575, 0.987,

0.888, 0.510, 0.391, 0.245, 0.146, 0.037, 0.009, -0.059, -0.097 -0.183, and -0.219. The sum of the SMC's was 8.207, and the five greatest latent roots above accounted for this sum. Accordingly, five factors were determined.

The decision rule was to reject hypothesis 2a for tests classified in a product category if they did not have "significant" loadings of .30 or greater in absolute value on at least one common factor. Otherwise, the hypothesis was not rejected.

Summaries of the test data and the results of the factor analysis are given in tables II and III. Table II contains the means, standard deviations, divergent productivity ratio and reliability estimates of the tests. The divergent productivity ratio (P) was determined as the ratio of the number of students who produced at least two responses to the total number of students in the sample, and the reliability estimate (R) was the calculated communality of the tests. Table II also provides a guide to the variables referred to in table III.

The varimax solution from an initial principal axis solution for the facility measures of main sample 1 is given in table 4. The five factors obtained were symbolized by F_1, F_2, F_3, F_4, and F_5, as in Table III.

TABLE II

MEANS, STANDARD DEVIATIONS, DIVERGENT PRODUCTIVITY RATIOS AND RELIABILITY ESTIMATES FOR MEASURES OF FACILITY IN PRODUCTION OF STUDENTS IN MAIN SAMPLE 1

(NUMBER OF STUDENTS (N) = 40)

Symbolic Description of variable	Test No.[1]	Description	Mean	SD	P	R
DMaU(F) 1	1	Facility in DMaU	6.85	3.35	0.98	0.82
DMaU(F) 2	I	Facility in DMaU	5.50	3.11	0.95	0.68
DMaS(F) 1	IV	Facility in DMaS	0.78	0.82	0.15	0.58
DMaS(F) 2	4	Facility in DMaS	1.78	1.54	0.48	0.59
DMaT(F) 1	5B	Facility in DMaT	2.75	2.42	0.63	0.43
DMaT(F) 2	5	Facility in DMaT	1.03	0.96	0.28	0.75
DMaI(F) 1	VI	Facility in DMaI	2.68	2.02	0.80	0.64
DMaI(F) 2	6B	Facility in DMaI	2.05	1.76	0.50	0.33
CPS 1	VII	Convergent Problem Solving	5.08	2.03		
CPS 2	VIIB	CPS	5.28	1.47		
LTV		Lorge-Thorndike Verbal I.Q. Test[2]	130.53	11.15		0.68
LTNV		L-T Non-Verbal[2] I.Q. Test	129.70	9.75		0.56
SMM1		Subject Matter Mastery[3]	20.74	5.55		0.74

[1] The test numbers refer to test numbers as in section 4.2-2.

[2] Here N = 32. The other I.Q. scores were missing

[3] Here N = 35. The other marks were missing.

TABLE III

VARIMAX ROTATED PRINCIPAL FACTOR SOLUTION FOR THIRTEEN VARIABLES INCLUDING EIGHT MEASURES OF FACILITY IN DIVERGENT PRODUCTION[1]

Communality estimates: SMC's

	Common Factors					Communalities	
Variable	F_1	F_2	F_3	F_4	F_5	Original	Calculated
DMaU(F) 1	-	72	-38	-	-37	80	82
DMaU(F) 2	-	30	-	-	-66	65	68
DMaS(F) 1	55	-	-	-	-39	59	58
DMaS(F) 2	70	-	-	-	-	58	59
DMaT(F) 1	-	48	-	-	-33	48	43
DMaT(F) 2	76	-	-33	-	-	70	75
DMaI(F) 1	-	-	-73	-	-	60	64
DMaI(F) 2	-	-	-53	-	-	34	33
CPS 1	-	45	-	-70	-	76	78
CPS 2	46	-	-	-71	-	79	79
LTV	35	56	-42	-	-	65	68
LTNV	-	69	-	-	-	55	56
SMM1	73	36	-	-	-	76	74

[1]Only loadings of .30 and greater in absolute value are included. Decimal points have been omitted

Hypothesis 2a was not rejected for the test measures of facility in the divergent production of mathematical units. Both DMaU(F) 1, and DMaU(F) 2, loaded "significantly" on factors F_2 and F_5.

Hypothesis 2a was also not rejected for the test measures of facility in the divergent production of mathematical systems. Both DMaS(F) 1 and 2, loaded "significantly" on factor F_1.

Hypothesis 2a was rejected for the test measures of facility in the divergent production of mathematical transformations. DMaT(F) 1 and 2 did not load "significantly" on any one factor.

Hypothesis 2a was not rejected for the test measures of facility in the production of mathematical implications. Both DMaI(F) 1 and 2, loaded significantly on factor F_3.

<u>6.3-3 Discussion - Facility Measures of Main Sample 1</u>

The analysis tended to support the hypothesis that test measures of facility in DMaU, DMaS, and DMaI, were "suitably classified." Measures of facility in DMaT were not considered "suitably classified" on the basis of the analysis.

The decision was made to retain DMaT(F) 2 as the measure of facility in the divergent production of mathematical transformations, and to consider DMaT (F) 1 as not "suitably classified." The DMaT(F) 2 measure had the higher reliability estimate of 0.75 as against that of DMaT(F) 1 with 0.43. Further, DMaT(F) 1 measure had much in common with the units measures, since its two "significant" loadings were on factors F_2 and F_5, the factors which the units test measures determined. On the other hand, the DMaT(F) 2 measure loaded "significantly" on factors F_1 and F_3, factors on which the systems and

implications test measures, respectively, loaded significantly. It was considered that transformations was a rather complex operation, and more likely to be identified with systems and implications, than with units.

6.3-3-1 Identification of Factors

The factors were identified in terms of the test measures which determine the factor and the highest loadings of the reference measures. The following names were thus suggested for the factors.

Factor	Identification
F_1	Subject Mastery and Facility in Systems.
F_2	Intelligence and Facility in Units.
F_3	Facility in Implications.
F_4	Convergent Problem Solving.
F_5	Facility in Units.

6.3-4 Analysis and Results -- Facility Measures of Main Sample 2

Facility measures for main sample 2 were obtained in seven DP tests, four in the DMaU category, and three in the DMaS category. The analysis included two CPS test measures, California Test of Mental Maturity (CTMM) intelligence test measures, and a subject mastery test (SMM2). A varimax solution was obtained from an initial principal axis solution of the matrix of intercorrelations of the eleven variables with squared multiple correlations in the diagonal. The eleven latent roots obtained were: 3.067, 1.455, 0.630, 0.283, 0.129, 0.101, 0.136, -0.116, -0.135, -0.195, and -0.282. The sum of the SMC's was 4.951, and the three greatest latent roots accounted for this sum. Accordingly three factors

were determined.

The decision rule was to reject hypothesis 2a for tests classified in a product category if they did not have "significant" loadings of .30 or greater in absolute value on at least one factor. Otherwise the hypothesis was rejected.

Summaries of the test data and the results of the factor analysis are given in tables IV and V. The varimax solution from an initial principal axis solution for the facility measures of main sample 2 is given in table V. The three factors obtained were symbolized by F_1, F_2, and F_3.

Hypothesis 2a was rejected for the four DMaU(F) measures, since the four did not load "significantly" on any one factor. However, since DMaU(F) 1, 2, and 3 loaded "significantly" on factor F_1, the hypothesis was not rejected for the three DMaU(F) 1, 2, and 3 measures.

Hypothesis 2a was rejected for the three DMaS(F) measures. since they did not load "significantly" on any one factor. It was observed, however, that DMaS(F) 2 did not load "significantly" on any factor, but that it had its highest loading of 0.27 on F_3, as did DMaS(F) 1.

6.3-5 Discussion -- Facility Measures of Main Sample 2

The analysis indicated that three of four DMaU(F) test measures were suitably classified. The DMaS(F) test measures were not suitably classified, but two of the three DMaS(F) tests had their highest loadings on the same factor F_3, and were accepted as "suitably classified."

TABLE IV

MEANS, STANDARD DEVIATIONS, DIVERGENT PRODUCTIVITY RATIO AND RELIABILITY ESTIMATES FOR MEASURES OF FACILITY IN PRODUCTION OF STUDENTS IN MAIN SAMPLE 2

(NUMBER OF STUDENTS (N) = 62)

Symbolic Description of variable	Test No.[1]	Description	Mean	SD	P	R
DMaU(F) 1	1	Facility in DMaU	4.85	3.08	0.92	0.60
DMaU(F) 2	I	Facility in DMaU	4.34	2.67	0.87	0.68
DMaU(F) 3	IB	Facility in DMaU	2.98	1.65	0.81	0.34
DMaU(F) 4	1B	Facility in DMaU	2.29	1.32	0.69	0.28
DMaS(F) 1	IV	Facility in DMaS	0.53	0.59	0.05	0.30
DMaS(F) 2	4	Facility in DMaS	1.18	0.99	0.29	0.08
DMaS(F) 3	4B	Facility in DMaS	3.29	1.50	0.95	0.30
CPS 2	VII	Convergent Problem Solving	3.39	1.67		0.76
CPS 3	VIIB	Convergent Problem Solving	3.40	1.76		0.83
CTMM		California Test of Mental Maturity[2]	114.98	12.26		0.39
SMM 2		Subject matter[3] mastery	19.51	4.61		0.57

[1]The test numbers are as in section 4.2-2

[2]Here N = 47. The other CTMM scores were missing

[3]Here N = 59. The other marks were missing.

TABLE V

VARIMAX ROTATED PRINCIPAL FACTOR SOLUTION FOR ELEVEN VARIABLES INCLUDING SEVEN MEASURES OF FACILITY IN DIVERGENT PRODUCTION[1] -- MAIN SAMPLE 2

Communality estimates: SMC's

	Common Factors			Communalities	
Variable	F_1	F_2	F_3	Original	Calculated
DMaU(F) 1	70	-	-34	59	60
DMaU(F) 2	80	-	-	61	68
DMaU(F) 3	43	-	-31	37	34
DMaU(F) 4	-	-	-44	28	28
DMaS(F) 1	-	-	-49	29	30
DMaS(F) 2	-	-	-27	14	08
DMaS(F) 3	55	-	-	33	30
CPS 2	-	-85	-	73	76
CPS 3	-	-90	-	77	83
CTMM	30	-	-49	37	39
SMM	-	-	-69	49	57

[1]Only loadings of .27 and greater in absolute value are included. Decimal points have been omitted.

6.3-5-1 Identification of Factors.

The factors were identified in terms of the test measures which determined the factor and the highest loadings of the reference measures. The following names were thus suggested for the factors.

Factor	Identification
F_1	Facility in Units.
F_2	Convergent Problem Solving.
F_3	Intelligence, Subject Mastery, and Facility in Systems.

6.3-6 Factor Correspondences

Similarities were observed between the factor structures of the facility measures of main samples 1 and 2, and correspondences may be made as follows: Factor F_1 of main sample 2 corresponds to factor F_5 of main sample 1 as factors of facility in units. Factor F_2 of main sample 2 corresponds to factor F_4 of main sample 1 as convergent problem solving factors, and factor F_3 corresponds roughly to factor F_1 of main sample 1 as indicating subject mastery and facility in systems.

The observed correspondences tend to indicate that the factor constructs underlying the facility measures are stable over the two samples. This is evidence supporting the validity of the tests as measuring generalizable constructs.

6.3-7 Analysis and Results -- Variety Measures of Main Sample 1

Variety measures for main sample 1 were obtained in ten DP tests, two in each of the product categories, DMaU, DMaC, DMaS, DMaT, and DMaI. The analysis also included the two CPS test measures,

A varimax solution was obtained from an initial principal axis solution of the matrix of intercorrelations of the fifteen test variables, with SMC's as communality estimates. The fifteen latent roots were: 6.101, 1.627, 1.082, 0.715, 0.500, 0.416, 0.194, 0.140, 0.116, 0.032, -0.051, -0.101, -0.146, -0.166, and -0.0183. The sum of the SMC's was 10.274, and the greatest six roots accounted for this sum. Accordingly six principal factors were determined, accounting for the trace of the reduced correlation matrix.

The decision rule was to reject hypothesis 2b for test measures in a product category, if they did not have "significant" leadings of .30 or greater in absolute value on at least one common factor. Otherwise, the hypothesis was not rejected.

Summaries of the test data and the results of the factor analysis are given in tables VI and VII. The varimax solution for the variety measures of main sample 1 is given in table VI. The six factors were symbolized F_1, F_2, F_3, F_4, F_5, and F_6, as in table VII.

Hypothesis 2b was not rejected for the test measures of divergent production of variety in mathematical units, since both DMaU(V) 1 and 2, loaded "significantly" on factor F_1.

Hypothesis 2b was also not rejected for the test measures of divergent production of variety in mathematical classes, since both DMaC(V) 1 and 2, loaded "significantly" on factor F_3.

Hypothesis 2b was also not rejected for the test measures of divergent production of mathematical systems, since both DMaS(V) 1 and 2, loaded "significantly" on factor F_2.

TABLE VI

MEANS, STANDARD DEVIATIONS, DIVERGENT PRODUCTIVITY RATIOS AND RELIABILITY (COMMUNALITY) ESTIMATES FOR MEASURES OF VARIETY IN PRODUCTION OF STUDENTS IN MAIN SAMPLE 1

(Number of Students (N) = 40)

Symbolic Description of variable	Test No.[1]	Description	Mean	SD	R
DMaU(V) 1	1	Variety in DMaU	2.70	1.27	0.54
DMaU(V) 2	I	Variety in DMaU	2.75	1.26	0.49
DMaC(V) 1	II	Variety in DMaC	1.73	0.95	0.80
DMaC(V) 2	2	Variety in DMaC	1.55	0.67	0.59
DMaS(V) 1	IV	Variety in DMaS	0.20	0.51	0.93
DMaS(V) 2	4	Variety in DMaS	0.70	1.10	0.64
DMaT(V) 1	5B	Variety in DMaT	1.85	1.78	0.77
DMaT(V) 2	5	Variety in DMaT	0.30	0.60	0.86
DMaI(V) 1	VI	Variety in DMaI	0.63	1.02	0.58
DMaI(V) 2	6B	Variety in DMaI	0.75	1.13	0.68
CPS 1	VII	C. Prob. Solving	5.08	2.03	0.75
CPS 2	VIIB	C. Prob. Solving	5.28	1.47	0.86
LTV		Lorge-Thorndike Verbal I.Q. Test[2]	130.53	11.15	0.73
LTNV		Lorge-Thorndike Non-Verbal I.Q. Test[2]	129.70	9.75	0.52
SMM 1		Subject matter mastery[3]	20.74	5.55	0.71

[1] The test numbers are as in section 4.2-2

[2] Here N = 32. The other I.Q.'s were missing.

[3] Here N = 35. The other marks were missing.

TABLE VII

VARIMAX ROTATED PRINCIPAL FACTOR SOLUTION FOR FIFTEEN VARIABLES INCLUDING TEN MEASURES OF VARIETY IN DIVERGENT PRODUCTION[1]
MAIN SAMPLE 1

Communality estimates: SMC's

	Common Factors						Communalities	
Variable	F_1	F_2	F_3	F_4	F_5	F_6	Original	Calculated
DMaU(V) 1	43	-	-	58	-	-	52	54
DMaU(V) 2	68	-	-	-	-	-	50	49
DMaC(V) 1	54	-	-42	-	-32	-40	73	80
DMaC(V) 2	-	-	-74	-	-	-	55	59
DMaS(V) 1	43	-82	-	-	-	-	88	93
DMaS(V) 2	-	-54	-42	38	-	-	64	64
DMaT(V) 1	58	-45	-	43	-	-	79	77
DMaT(V) 2	-	-82	-	36	-	-	83	86
DMaI(V) 1	-	-74	-	-	-	-	63	58
DMaI(V) 2	-	-54	-	-	-	-52	63	68
CPS 1	33	-	-	38	-66	-	74	75
CPS 2	-	-34	-33	-	-75	-	82	86
LTV	-	-	-	66	-	-35	74	73
LTNV	50	-	-31	-	-	-	56	52
SMM 1	-	-39	-	61	-35	-	72	71

[1]Only loadings of .30 and greater in absolute value are included. Decimal points have been omitted.

Hypothesis 2b was also not rejected for the test measures of divergent production of variety in mathematical transformations, since DMaT(V) 1 and 2, loaded "significantly" on factor F_2.

Hypothesis 2b was also not rejected for test measures of divergent production of variety in mathematical implications, since DMaI(V) 1 and 2, loaded "significantly" on Factor F_2.

6.3-8 Discussion -- Variety Measures of Main Sample 1

On the basis of the analysis, the variety measures in each of the product categories tested in main sample 1, were considered as "suitably classified."

6.3-8-1 Identification of Factors.

The factors were identified in terms of the test measures which determined the factor, and the highest loadings of the reference measures. The following names were suggested for the factors.

Factor	Identification
F_1	Non-Verbal Intelligence and Variety in Units.
F_2	Variety in Systems, Transformations, and Implications
F_3	Variety in Classes
F_4	Subject Mastery, Verbal Intelligence, and Variety in transformations.
F_5	Convergent Problem Solving.
F_6	--------

No name is suggested for factor F_6, since it was not determined by tests in any one product category, and it did not contain the highest loading of any of the reference measures.

6.3-9 Analysis and results for variety measures of Main Sample 2

Variety measures for main sample 2 were obtained in ten DP tests, four in DMaU category, three in DMaC category, and three in DMaS category. It was observed that each student scored zero in the DMaS(V) 1 measure of test IV. Accordingly, only nine DP measures were used in the analysis. The analysis also included two CPS test measures. California Test of Mental Maturity (CTMM) measures, and a subject mastery test. A varimax solution was obtained from an initial principal factor solution of the matrix of intercorrelations of the thirteen variables with squared multiple correlations in the diagonal The thirteen latent roots obtained were: 2.986, 1.142, 0.736, 0.611, 0.338, 0.302, 0.140, 0.009, -0.110, -0.144, -0.179, -0.260, and -0.323. The sum of the SMC's was 5.247, and the four greatest latent roots accounted for this sum. Accordingly, four principal factors were determined.

The decision rule was to reject the hypothesis for tests classified in a product category, if they did not have "significant loadings on at least one common factor. A "significant" loading was considered as a loading which was equivalent to .30 or greater in absolute value.

Summaries of the data and the results of the factor analysis are given in tables VIII and IX. The varimax solution for variety measures of main sample 2 is given in table IX. The factors were symbolized as F_1, F_2, F_3, and F_4, as in table IX.

Hypothesis 2b was rejected for the four test measures of divergent production of variety in mathematical units. However, it was observed that DMaU(V) 1 and 2 loaded "significantly" on factor

TABLE VIII

MEANS, STANDARD DEVIATIONS, DIVERGENT PRODUCTIVITY RATIOS AND RELIABILITY (COMMUNALITY) ESTIMATES FOR MEASURES OF VARIETY IN PRODUCTION OF STUDENTS IN MAIN SAMPLE 2

Number of Students (N) = 62

Symbolic Description of variable	Test No.[1]	Description	Mean	SD	R
DMaU(V) 1	1	Variety in DMuU	2.53	1.40	0.44
DMaU(V) 2	I	Variety in DMaU	2.19	1.23	0.38
DMaU(V) 3	IB	Variety in DMaU	1.90	1.09	0.55
DMaU(V) 4	1B	Variety in DMaU	1.40	0.75	0.19
DMaC(V) 1	II	Variety in DMaC	1.19	1.09	0.26
DMaC(V) 2	2	Variety in DMaC	1.55	0.76	0.29
DMaC(V) 3	IIB	Variety in DMaC	0.55	0.71	0.36
DMaS(V) 2	4	Variety in DMaS	0.19	0.50	0.17
DMaS(V) 3	4B	Variety in DMaS	2.73	1.23	0.27
CPS 1	VII	C. Prob. Solving	3.39	1.67	0.79
CPS 3	VIIB	C. Prob. Solving	3.40	1.76	0.82
CTMM		California Test[2] of Mental Maturity	114.98	12.26	0.36
SMM 2		Subject matter[3] mastery	19.51	4.61	0.59

[1]The test numbers are as in section 4.2-2

[2]Here N = 47. The other CTMM scores were missing.

[3]Here N = 59. The other marks were missing.

TABLE IX

VARIMAX ROTATED PRINCIPAL FACTOR SOLUTION FOR THIRTEEN VARIABLES INCLUDING NINE MEASURES OF VARIETY IN DIVERGENT PRODUCTION[1] IN STUDENTS IN MAIN SAMPLE 2

Communality estimates: SMC's

	Common Factors				Communalities	
Variable	F_1	F_2	F_3	F_4	Original	Calculated
DMaU(V) 1	-	-	-	63	37	44
DMaU(V) 2	-	-	-	56	35	38
DMaU(V) 3	60	-	34	-	52	55
DMaU(V) 4	39	-	-	-	22	19
DMaC(V) 1	-	-	-50	-	21	26
DMaC(V) 2	-	-	-51	-	25	29
DMaC(V) 3	47	-	-34	-	37	36
DMaS(V) 2	-	-	-	37	19	17
DMaS(V) 3	49	-	-	-	34	27
CPS 2	-	-85	-	-	76	79
CPS 3	-	-87	-	-	77	82
CTMM	51	-	-	-	34	36
SMM 2	67	-	-	32	55	59

[1]Only loadings of .30 and greater in absolute value are included. Decimal points have been omitted.

F_4, and DMaU(V) 3 and 4 loaded "significantly" on factor F_1.

Hypothesis 2b was not rejected for the DMaS(V) measures, since the three measures loaded "significantly" on factor F_3.

The hypothesis was rejected for the DMaS(V) measures, since the two measures loaded on different factors.

6.3-10 Discussion -- Variety Measures of Main Sample 2

On the basis of the analysis, the four DMaU(V) measures were considered to measure two aspects of variety in the production of mathematical units. DMaU(V) 1 and 2 were considered as "suitably classified" to measure one aspect, and DMaU(V) 3 and 4 were considered as suitably classified to measure the other aspect.

The variety measures of divergent production of mathematical classes were considered as "suitably classified" on the basis of the analysis.

The measures of divergent production of mathematical systems were considered to have not been suitably classified. Their individual reliabilities of .17 and .27 were considered as too low for any of them to be considered independently as suitable measures of a construct. It should be noted here that one of the three measures of divergent production of mathematical systems was not included in the analysis, because every subject's variety score was zero on that test. It was concluded that in general the DMaS tests were unsuitable as measures of variety in the production of mathematical systems in main sample 2.

6.3-10-1 Identification of Factors

The factors were identified in terms of the test measures which determined the factor, and the highest leadings of the

reference measures. The following names were suggested for the factors.

Factor	Identification
F_1	Intelligence, Subject Mastery, and Variety Units.
F_2	Convergent Problem Solving
F_3	Variety in Classes.
F_4	Variety in Units.

6.3-11 Factor Correspondences -- Variety Measures

Similarities were observed between the factor structures of the variety measures of main samples 1 and 2, and correspondences may be made as follows: Factor F_1 of main sample 2 corresponds to factor F_1 of main sample 1 as factors of intelligence and variety in units. Factor F_2 of main sample 2 corresponds to factor F_5 of main sample 1 as convergent problem solving factors. Factor F_3 of main sample 2 corresponds to factor F_3 of main sample 1, as factors of variety in classes, and Factor F_4 of main sample 2 corresponds to factor F_1 of main sample 1. Since factors F_1 and F_4 of main sample 2 appear to correspond to the same factor F_1 of main sample 1, this suggests that some linear combination of factors F_1 and F_4 of main sample 2 would result in a suitable single factor corresponding to factor F_1 of main sample 1.

The observed correspondences indicate that the factor constructs underlying the variety measures for which comparable evidence is available, are stable over the two samples. This strong evidence supporting the validity of the tests as measuring generalizable constructs.

6.3-12 Analysis and Results -- Novelty Measures of Main Sample 1

Novelty measures for main sample 1 were obtained in ten DP tests, two in each of the product categories, DMaU, DMaC, DMaS, DMaT, and DMaI. The analysis also included two tests of convergent problem solving, Lorge-Thorndike verbal and non-verbal intelligence test measures, and a subject mastery test.

A varimax solution of the reduced correlation matrix of the fifteen variables was obtained. The fifteen latent roots obtained were: 5.711, 0.993, 0.819, 0.773, 0.552, 0.510, 0.415, 0.193, 0.069, 0.026, -0.032, -0.133, -0.171, -0.181, and -0.278. The sum of the SMC's was 9.266, and the greatest six roots accounted for this sum. Accordingly six principal factors were determined, accounting for the trace of the reduced correlation matrix.

The decision rule was to reject hypothesis 2c for tests classified in a product category if they did not have "significant" loadings of .30 or greater in absolute value on at least one common factor.

Summaries of the data and the results of the factor analysis are given in tables X and XI. The varimax solution for novelty measures of main sample 1 is given in table XI. The factors were symbolized as F_1, F_2, F_3, F_4, F_5, and F_6.

Hypothesis 2c was not rejected for test measures of novelty in the production of mathematical units. Both DMaU(N) 1 and 2 loaded "significantly" on factor F_5.

Hypothesis 2c was also not rejected for test measures of novelty in the divergent production of mathematical classes. Both DMaC(N) 1 and 2 loaded "significantly" on factor F_4.

TABLE X

MEANS, STANDARD DEVIATIONS AND RELIABILITY (COMMUNALITY) ESTIMATES FOR MEASURES OF NOVELTY IN PRODUCTION OF STUDENTS IN MAIN SAMPLE 1

Number of Students (N) = 40

Symbolic Description of variable	Test No.[1]	Description	Mean	SD	R
DMaU(N) 1	1	Novelty in DMaU	6.10	1.80	0.37
DMaU(N) 2	I	Novelty in DMaU	5.98	1.71	0.55
DMaC(N) 1	II	Novelty in DMaC	1.10	1.34	0.82
DMaC(N) 2	2	Novelty in DMaC	1.20	1.33	0.62
DMaS(N) 1	IV	Novelty in DMaS	1.88	2.05	0.79
DMaS(N) 2	4	Novelty in DMaS	3.85	3.16	0.65
DMaT(N) 1	5B	Novelty in DMaT	3.78	2.97	0.43
DMaT(N) 2	5	Novelty in DMaT	1.75	2.20	0.77
DMaI(N) 1	VI	Novelty in DMaI	3.23	2.59	0.40
DMaI(N) 2	6B	Novelty in DMaI	4.13	2.79	0.62
CPS 1	VII	Convergent Problem Solving	5.08	2.03	0.77
CPS 2	VIIB	Convergent	5.28	1.47	0.74
LTV		L-T Verbal I.Q.[2]	130.53	11.15	0.68
LTNV		L-T Non-Verbal[2]	129.70	9.75	0.49
SMM 1		Subject Mastery[3]	20.74	5.55	0.67

[1] The test numbers are as in section 4.2-2

[2] Here N = 32. The other I.Q.'s were missing.

[3] Here N = 35. The other marks were missing.

TABLE XI

VARIMAX ROTATED PRINCIPAL FACTOR SOLUTION FOR FIFTEEN VARIABLES INCLUDING TEN MEASURES OF NOVELTY IN PRODUCTION[1]
MAIN SAMPLE 1

Communality estimates: SMC's

	Common Factors						Communalities	
Variables	F_1	F_2	F_3	F_4	F_5	F_6	Original	Calculated
DMaU(N) 1	-	-	-	-	50	-	53	37
DMaU(N) 2	-	-	-	-	66	-	50	55
DMaC(N) 1	-	38	-40	-52	42	-	78	82
DMaC(N) 2	-	-	-	-76	-	-	58	62
DMaS(N) 1	36	34	-	-	30	-65	74	79
DMaS(N) 2	75	-	-	-	-	-	60	65
DMaT(N) 1	-	31	-	-	48	-	48	43
DMaT(N) 2	70	35	-	-	-	-32	74	77
DMaI(N) 1	-	-	-39	-	-	-48	41	40
DMaI(N) 2	-	-	-75	-	-	-	56	62
CPS 1	-	79	-	-	31	-	74	77
CPS 2	-	73	-	-	-	-	73	74
LTV	35	42	-46	-	38	-	67	68
LTNV	-	32	-	-	50	-	53	49
SMM 1	53	46	-	-	31	-	69	67

[1]Only loadings of .30 and greater in absolute value are included. Decimal points have been omitted.

Hypothesis 2c was also not rejected for test measures of novelty in the divergent production of mathematical systems. Both DMaS(N) 1 and 2 loaded "significantly" on factor F_1.

Hypothesis 2c was also not rejected for test measures of novelty in the divergent production of mathematical transformations. Both DMaS(N) 1 and 2 loaded "significantly" on factor F_6.

6.3-13 Discussion -- Novelty Measures of Main Sample 1

On the basis of the analysis, the novelty measures in each of the product categories tested in main sample 1, were considered as "suitably classified."

6.3-13-1 Identification of Factors

The factors were identified in terms of the test measures which determined the factor and the highest loadings of the reference measures. The following names were thus suggested for the factors.

Factor	Identification
F_1	Subject mastery and novelty in systems.
F_2	Intelligence, Convergent Problem Solving, and Novelty in Transformations.
F_3	Verbal Intelligence and Novelty in Implications.
F_4	Novelty in Classes.
F_5	Intelligence and Novelty in Units.
F_6	Novelty in Transformations.

6.3-14 Analysis and Results -- Novelty Measures of Main Sample 2

Novelty measures for main sample 2 were obtained in ten DP tests, four in the DMaU category, three in the DMaC category, and three in the DMaS category. The analysis also included two tests of

convergent problem solving, California Test of Mental Maturity, and a subject mastery test.

A varimax solution of the reduced correlation matrix (with squared multiple correlations in the leading diagonal) was obtained. The fourteen latent roots were: 2.872, 1.248, 0.940, 0.698, 0.511, 0.399, 0.269, 0.196, 0.028, -0.059, -0.129, -0.226, -0.289, and -0.323. The sum of the SMC's was 6.135, and the greatest five factors accounted for this sum. Accordingly, five principal factors were determined.

The decision rule was to reject hypothesis 2c for tests classified in a product category if they did not have "significant" loadings of .30 or greater in absolute value on at least one factor.

Summaries of the data and the results of the factor analysis are given in tables XII and XIII. The varimax solution for the novelty measures of main sample 2 is given in table XIII. The factors were symbolized as F_1, F_2, F_3, F_4, and F_5.

Hypothesis 2c was rejected for the four DMaU(N) measures, since the four did not load "significantly" on any one factor. However, since DMaU(N) 1, 2, and 3, loaded "significantly" on factor F_3, the hypothesis was not rejected for the three DMaU(N) 1, 2, and 3 measures. Furthermore, since DMaU(N) 1 and 4, loaded "significantly" on factor F_2, the hypothesis was not rejected for DMaU(N) 1 and 4. Hypothesis 2c was rejected for the three DMa(C(N) measures, since they did not load "significantly" on any one factor. However, the hypothesis was not rejected for the two DMaU(N) 1 and 2 measures, since they loaded "significantly" on factor F_4.

TABLE XII

MEANS, STANDARD DEVIATIONS AND LOWER BOUND RELIABILITY ESTIMATES FOR MEASURES OF NOVELTY IN PRODUCTION OF STUDENTS IN MAIN SAMPLE 2

Number of Students (N) = 62

Symbolic Description of variable	Test No.[1]	Description	Mean	SD	R
DMaU(N) 1	1	Novelty in DMaU	4.89	2.27	0.47
DMaU(N) 2	I	Novelty in DMaU	5.27	2.06	0.51
DMaU(N) 3	IB	Novelty in DMaU	4.76	2.36	0.53
DMaU(N) 4	1B	Novelty in DMaU	3.44	2.33	0.32
DMaC(N) 1	II	Novelty in DMaC	1.05	0.89	0.22
DMaC(N) 2	2	Novelty in DMaC	1.58	1.94	0.26
DMaC(N) 3	IIB	Novelty in DMaC	0.94	1.28	0.33
DMaS(N) 1	IV	Novelty in DMaS	0.97	1.00	0.25
DMaS(N) 2	4	Novelty in DMaS	2.18	2.39	0.22
DMaS(N) 3	4B	Novelty in DMaS	5.34	1.65	0.32
CPS 1	VII	Convergent Problem Solving	3.39	1.67	0.80
CPS 3	VIIB	Convergent Problem Solving	3.40	1.76	0.82
CTMM		C. Test of Mental Maturity[2]	114.98	12.26	0.50
SMM 2		Subject Mastery[3]	19.51	4.61	0.71

[1]The test numbers are as in section 4.2-2

[2]Here N = 47. The other CTMM scores were missing.

[3]Here N = 59. The other marks were missing.

TABLE XIII

VARIMAX ROTATED FACTOR SOLUTION FOR FOURTEEN VARIABLES INCLUDING TEN MEASURES OF NOVELTY IN PRODUCTION OF STUDENTS[1] IN MAIN SAMPLE 2

Communality estimates: SMC's

	Common Factors					Communalities	
Variables	F_1	F_2	F_3	F_4	F_5	Original	Calculated
DMaU(N) 1	-	36	56	-	-	44	47
DMaU(N) 2	-	-	71	-	-	44	51
DMaU(N) 3	-	-	53	-	-43	54	53
DMaU(N) 4	-	56	-	-	-	28	32
DMaC(N) 1	-	-	-	-34	-	21	22
DMaC(N) 2	-	-	-	-49	-	21	26
DMaC(N) 3	-	54	-	-	-	35	33
DMaS(N) 1	33	-	-	-	-	35	26
DMaS(N) 2	-	-	-	-42	-	29	22
DMaS(N) 3	-	-	-	-	-53	32	32
CPS 1	88	-	-	-	-	75	80
CPS 3	88	-	-	-	-	77	82
CTMM	-	-	-	-	-60	51	50
SMM 2	-	-	38	-34	-61	68	71

[1]Only loadings of .30 and greater in absolute value are included. Decimal points have been omitted.

Hypothesis 2c was rejected for the three DMaS(N) measures. Each loaded "significantly" on a different factor.

6.3-14 Discussion -- Novelty Measures of Main Sample 2

On the basis of the analysis, the four DMaU(N) measures were considered to be measuring two aspects of novelty in the divergent production of mathematical units. The three measures, DMaU(N) 1, 2, and 3, were considered to be "suitably classified" in determining one aspect of the ability, and the two measures, DMaU(N) 1 and 4 were considered to be "suitably classified" in determining the other aspect of the ability. It was considered that where it was desirable to have independent measures of the two aspects novelty in the divergent production of mathematical units, then the common test measure, DMaU(N) 1 should be associated with factor F_3, on which it had its highest loading.

The two measures of novelty in the divergent production of mathematical classes, DMaC(N) 1 and 2 were considered as "suitably classified" on the basis of the analysis.

On the basis of the analysis, the three measures of novelty in the production of systems were considered as not suitably classified. The reliability estimates of the three tests, DMaS(N) 1 2 and 3 were 0.26, 0.22 and 0.32 respectively, and these were considered too low for the measures to be taken as independent determinants of constructs.

6.3-14-1 Identification of Factors

The factors were identified in terms of the test measures

which determined the factor, and the highest loadings of the reference measures. The following names were suggested for the factors.

Factor	Identification
F_1	Convergent Problem Solving.
F_2	Novelty in Units.
F_3	Novelty in Units.
F_4	Novelty in Classes.
F_5	Intelligence and Subject Mastery.

6.3-15 Factor Correspondences - Novelty Measures.

Some similarities were observed between the factors of main sample 1 and 2, and correspondences may be made as follows: Factor F_1 of main sample 2 corresponds to factor F_2 of main sample 1, since they are factors of convergent problem solving. Factor F_3 of main sample 2 correspond to factor F_5 of main sample 1, since they are factors of novelty in the divergent production of mathematical units. Factor F_4 of main sample 2 correspond to factor F_4 of main sample 1 as factors of novelty in the divergent production of mathematical classes.

The observed correspondences indicate that the factor constructs underlying the novelty measures of units and classes, are stable over the two samples. This is strong evidence supporting the validity of the relevant tests as measuring generalizable constructs.

6.3-16 General Discussion -- Hypotheses 2a, 2b, and 2c.

Tests were considered to be "suitably classified" when there was evidence on a factor analytic study that they measured

the same ability. This ability was characterized in two ways, according to product category, and according to DP-measure. The product categories were units, classes, systems, transformations, and implications, and the DP-measures were facility, variety, and novelty in production. Such tests were considered as "suitably classified" in the same product-measure category. Suitable tests, classified in the product-measure categories on the basis of the experimental studies in main samples 1 and 2 are reported in tables XIV and XV.

It was found that all ten DP tests tried out in main sample 1, were suitable for some product-measure ability. It was also found that nine of the ten tests tried out in main sample 2 were suitable for some product-measure ability. The one test found completely unsuitable was the DMaS 3 test (test 4B). This indicates that all the tests of main sample 1, and nine out of ten tests of main sample 2, have demonstrated construct validity in terms of determining constructs indicating product-measure abilities.

It was found that when comparisons were made between factor structures of the two samples studied, noteworthy similarities were found between the two samples in terms of the factors of convergent problem solving, facility in units, classes, and systems, variety in units, and classes, and novelty in units, and classes. These similarities indicate a stability of the underlying constructs over the two samples, and indicates that some of the tests are structurally measures of constructs which have some stability.

6.3-16 Reliability Estimates -- Tests in Product-Measure Categories and Tests of Convergent Problem Solving

Reliability estimates of composite tests were estimated

TABLE XIV

"SUITABLE" TESTS CLASSIFIED IN PRODUCE-MEASURE CATEGORIES ON BASIS OF FACTOR ANALYSES OF MEASURES OF MAIN SAMPLE 1

Product-Measure Category	Variables	Test Nos.	Reliability Estimate[1]
Facility in Mathematical Units (FMU)	DMaU(F) 1 and 2	1 and I	0.79
Facility in Mathematical Systems (FMS)	DMaS(F) 1 and 2	IV and 4	0.66
Facility in Mathematical Transformations	DMaT(F) 2	5	0.75[2]
Facility in Mathematical Implications (FMI)	DMaI(F) 1 and 2	VI and 6B	0.59
Variety in Math. Units (VMU)	DMaU(V) 1 and 2	1 and I	0.56
Variety in Math. Classes (VMC)	DMaC(V) 1 and 2	II and 2	0.57
Variety in Math. Systems (VMS)	DMaS(V) 1 and 2	IV and 4	0.63
Variety in Math. Trans (VMT)	DMaT(V) 1 and 2	5B and 5	0.68
Variety in Math. Imp. (VMI)	DMaI(V) 1 and 2	VI and 6B	0.65
Novelty in Math. Units (VMU)	DMaU(N) 1 and 2	1 and I	0.59
Novelty in Math. Classes (VMC)	DMaC(N) 1 and 2	II and 2	0.71
Novelty in Math. Systems (VMS)	DMaS(N) 1 and 2	IV and 4	0.46
Novelty in Math. Trans. (VMT)	DMaT(N) 1 and 2	5B and 5	0.56
Novelty in Math. Imp. (VMI)	DMaI(N) 1 and 2	VI and 6B	0.54

[1] This is the reliability estimate for the mean of the standard scores of the measures, using analysis of variance techniques (Winer, 1962, p. 124-132).

[2] The estimate in this case was the communality estimate reported in table III.

TABLE XV

"SUITABLE" TESTS CLASSIFIED IN PRODUCT-MEASURE CATEGORIES ON BASIS OF FACTOR ANALYSES OF MEASURES OF MAIN SAMPLE 2

Product-Measure Category	Variables	Test No.	Reliability[1] Estimate
Facility in Mathematical Units (FMU)	DMaU(F) 1, 2 and 3.	1, I, and IB.	0.73
Facility in Mathematical Systems (FMS)	DMaS(F) 1 and 2	IV and 4	0.25
Variety in Math. Units 1 (VMU 1)	DMaU(V) 1 and 2	1 and I	0.70
Variety in Math. Units 2 (VMU 2)	DMaU(V) 3 and 4	IB and 1B.	0.31
Variety in Math. Classes (VMC)	DMaC(V) 1, 2 and 3	II, 2 and IIB.	0.45
Novelty in Math. Units (NMU)	DMaU(N) 1, 2, and 3	1, I and IB.	0.66
Novelty in Math. Classes (NMC)	DMaU(N) 1 and 2.	II and 2.	0.43

[1]This is the reliability estimate for the mean of the standard scores of the measures, using analysis of variance techniques. (Winer, 1962, p. 124-132.)

using analysis of variance techniques (Winer, 1962, p. 124-132) on standard scores of the tests. The reliability estimates reported were those of the mean of the measurements. They indicate internal consistency in determining a single construct.

The reliability estimates of composite tests in the product measure categories may be found in table XIV and XV. The reliability estimate for the composite CPS tests of main sample 1 was 0.84, and that for the CPS tests of main sample 2 was 0.91.

6.3-17 Hypotheses 3 and 4

Hypotheses 3 and 4 were tested to investigate expected relationships between divergent production abilities and abilities involved in convergent problem solving and subject mastery.

Hypothesis 3 That divergent production abilities in school mathematics predict school achievement in mathematics significantly.

Hypothesis 4 That divergent production abilities in mathematics predict convergent problem solving ability significantly.

6.3-18 Preliminary Discussion

The composite CPS tests were used as measures of problem solving ability and the subject mastery tests were used as measures of school achievement. The product-measures (see tables XIV and XV) were used as measures of DP abilities.

6.3-19 Analysis and Results -- Main Sample 1

Hypothesis 3 was tested in terms of facility, variety and novelty measures independently. The decision rule was to reject the hypothesis for a DP measure, if the multiple correlation of the predictor variables with the criterion measure was not significantly different from zero at the

.05 level, and not to reject the hypothesis otherwise. The test for significance used for the multiple correlation coefficient was from Fergusson (1966, p. 40). An F ratio was calculated, and the value of F was

$$F = (R^2/k) + [(1 - R^2)/(N - k - 1)],$$

Where R, N and k were the multiple correlation coefficient, number of observations, and number of predictors, respectively. The degree of freedom were k and (N - k - 1).

The simple product-moment correlation of each variable with regard to the two criteria of problem solving and achievement are given in table XVI. The results of the analysis of the significance of the multiple correlation using the DP variables as predictors are presented in table XVII.

It was found that each of the facility, variety, and novelty variables predicted problem solving ability significantly, the multiple correlation coefficients being 0.63, 0.69, and 0.67 respectively.

Hypothesis 4 was tested for main sample 1 in a similar way as described for the testing of hypothesis 3. The details of the testing may be found as for hypothesis 3 in tables XVI and XVII.

It was found that each of the facility, variety, and novelty variables predicted achievement significantly, the multiple correlation coefficients being 0.76, 0.68, and 0.75, respectively.

An inspection of the simple product moment correlations of the DP variables with each of the criteria independently, revealed that most of the abilities measured by these variables were significant predictors of convergent problem solving and achievement.

TABLE XVI

SIMPLE PRODUCT MOMENT CORRELATIONS OF DP PREDICTOR VARIABLES WITH CRITERION MEASURES FOR PROBLEM SOLVING AND ACHIEVEMENT MAIN SAMPLE 1

	PROBLEM SOLVING AS CRITERION[1] (N = 40)	ACHIEVEMENT AS CRITERION[2] (N = 35)
Variable	r	r
FM Units	0.45	0.50
FM Systems	0.59	0.72
FM Transformations	0.54	0.64
FM Implications	0.29	0.17
VM Units	0.39	0.40
VM Classes	0.55	0.29
VM Systems	0.54	0.61
VM Trans.	0.56	0.66
VM Implications	0.36	0.36
NMU	0.37	0.44
NMC	0.54	0.34
NMS	0.53	0.69
NMT	0.55	0.60
NMI	0.20	0.27

[1]For N = 40, correlations of .32 and greater are significant at the .05 level.

[2]For N = 35, correlations of .34 and greater are significant at the .05 level.

TABLE XVII

SUMMARY OF ANALYSIS OF SIGNIFICANCE OF MULTIPLE PREDICTION OF CONVERGENT PROBLEM SOLVING AND ACHIEVEMENT WITH DP VARIABLES AS PREDICTORS

Criterion	Predictors	Multiple R	F ratio	df_1	df_2	P	P(α)
Convergent Problem Solving	FMU, FMS, FMT, FMI.	0.63	6.17	4	35	.00	$<.05$
Achievement	FMU, FMS, FMT, FMI	0.76	10.67	4	30	.00	$<.05$
Convergent Problem Solving	VMU, VMC, VMS, VMT, VMI.	0.69	6.71	5	35	.00	$<.05$
Achievement	VMU, VMC, VMS, VMT, VMI.	0.68	5.47	5	29	.00	$<.05$
Convergent Problem Solving	NMU, NMC, NMS, NMT, NMI.	0.67	5.88	5	34	.00	$<.05$
Achievement	NMU, NMC NMS, NMT, NMI.	0.75	8.12	5	29	.00	$<.05$

6.3-20 Analysis and Results -- Main Sample 2

Hypotheses 3 and 4 were tested for main sample 2 in accordance with the procedures used in testing the hypotheses for main sample 1. Table XVIII gives the product-moment coefficient of each variable with the variables of each of the two criteria of problem solving and achievement in main sample 2. The results of the analysis of the significance of the multiple correlation using the DP variables as predictors are presented in table XIX

It was found that the facility measures predicted the convergent problem solving criterion significantly, at the .05 level of significance. The multiple correlation coefficient was 0.36.

It was also found that the variety and novelty measures did not significantly predict convergent problem solving. The multiple correlation coefficients of the variety and novelty measures with convergent problem solving were 0.29 and 0.23, respectively.

It was further found that each of the set of facility, variety, and novelty variables significantly predicted achievement, the multiple correlation coefficients being 0.61, 0.61, and 0.48, respectively.

It was also observed that most of the variables predicted achievement significantly, but none of the variables in variety and novelty predicted convergent problem solving significantly.

6.3-21 Discussion --- Hypotheses 3 and 4.

The findings that the DP - measure variables significantly relate to the achievement criterion, is of considerable importance in establishing one aspect of the construct validity of the tests.

TABLE XVIII

SIMPLE PRODUCT MOMENT CORRELATION COEFFICIENT OF DP PREDICTOR VARIABLES WITH CRITERION MEASURES FOR PROBLEM SOLVING AND ACHIEVEMENT
MAIN SAMPLE 2

	PROBLEM SOLVING AS CRITERION[1] (N = 62)	ACHIEVEMENT AS CRITERION[2] (N = 59)
Variable	r	r
FM Units	0.25	0.47
FM Systems	0.30	0.48
VM Units 1	0.21	0.34
VM Units 2	0.24	0.53
VM Classes	0.07	0.30
NMU	0.23	0.46
NM Classes	-0.03	0.11

[1]For N = 62, correlations of .25 and greater are significant at the .05 level.

[2]For N = 59, correlations of .26 and greater are significant at the .05 level

TABLE XIX

SUMMARY OF ANALYSIS OF SIGNIFICANCE OF MULTIPLE PREDICTION OF CONVERGENT PROBLEM SOLVING AND ACHIEVEMENT WITH DP VARIABLES AS PREDICTORS

Criterion	Predictors	Multiple R	F ratio	df_1	df_2	P	$P(\alpha)$
Convergent Problem Solving	FMU, FMS	0.36	4.43	2	59	.02	.05
Achievement	FMU, FMS	0.61	17.26	2	56	.00	.05
Convergent Problem Solving	VMU 1, VMU 2 VMC	0.29	1.78	3	58	.16	.05
Achievement	VMU 1, VMU 2 VMC	0.61	11.38	3	55	.00	.05
Convergent Problem Solving	NMU, NMC	0.23	1.64	2	59	.20	.05
Achievement	NMU, NMC	0.48	8.84	2	56	.00	.05

The tests were constructed to evoke inventive responses in students who were studying mathematics in the senior high school, and these responses were expected to be primarily based on the mathematics studied in school. The tests were subject-related, and it is therefore a critical expectation that these tests should be significantly related to the subject matter. It is thus a very important result that the test measures were found to relate significantly to achievement.

The finding that the DP measures predicted convergent problem solving significantly in main sample 1, and the facility measures of sample 2 predicted convergent problem solving significantly, but that the variety and novelty of main sample 2 measures did not predict convergent problem solving significantly indicates a need for further investigation. It may well be that while the connection between divergent production and convergent problem solving is fundamental in that they are both the result of combining ideas, yet certain concomitant variables are of crucial importance in establishing a significant connection between the two. These variables may be indicative of direction in thinking, evaluation, motivation, persistence, intuition, and sensitivity.

6.3-22 "Stable" tests and Hierarchical Investigations

Tests which were found to be "suitably classified" in their product categories for every DP measure obtained from them, were considered to have shown stability over the DP measures and were called "stable" tests. These tests were used in hierarchical investigations. They are described in table XX, and reliability estimates of the composite of the tests in their product-measure classifications

are given. These estimates were obtained by using analysis of variance procedures (Winer, 1962, p. 124-132) on the standard scores, a procedure which is equivalent to taking the average correlation of the variables as the average individual reliability of one measure, and using the Spearman-Brown formula to obtain the reliability estimate for the mean of all the measurements. The reliability estimates for the single transformation test were obtained from communality estimates reported in table III.

6.3-23 Hierarchical Investigations -- Preliminary Discussion

Investigations on hierarchical orderings were conducted using "stable" tests. The sequence of investigation was as follows:

(i) Investigation into the existence of a betweeness relationship among the variables and the establishment of order among the variables on the basis of this relationship.

(ii) Investigation that the order established in (i) above was hierarchical.

(iii) Investigation that observed betweeness relationships were indicative of complexity relationships.

Betweeness Relationships

Guttman (1954) has developed a notion of complexity among tests, such that a test t_2 is more complex than a t_1, if "it requires everything that t_1 does and more." If g is conceived as "the total complexity factor," Guttman's basic hypothesis is that for n tests, t_1, t_2, ... t_n, whose rank order of complexity was represented by their subscripts, the partial correlation $r_{jg.k} = 0$, $(j < k)$. Thus the correlation of a test t_j, with rank order j, and the total complexity factor g, with test t_k partialled out should be zero, whenever the rank order of t_j is less than the rank order of t_k. Guttman states that a set of tests whose intercorrelations satisfy

TABLE XX

"STABLE TESTS
TESTS WHICH WERE FOUND "SUITABLY CLASSIFIED" IN THEIR PRODUCT CATEGORIES FOR EVERY DP MEASURE OBTAINED FROM THEM

(a) MAIN SAMPLE 1

		Reliability estimates as measures of[1]		
Test No.	Classification	Facility	Variety	Novelty
1 and I	Units	0.79	0.56	0.59
2 and II	Classes	---	0.57	0.71
4 and IV	Systems	0.66	0.63	0.46
5	Transformations	0.75[2]	0.86[2]	0.77[2]
6B and VI	Implications	0.59	0.65	0.54

(b) MAIN SAMPLE 2

		Reliability estimates as measures of		
Test No.	Classification	Facility	Variety	Novelty
1 and I	Units	0.83	0.70	0.51
2 and II	Classes	---	0.57	0.43

[1]Reliability estimates were obtained for the mean of the standard scores of the measures, using analysis of variance techniques (Winer, 1962).

[2]The estimate in this case was the communality estimate reported in table III.

a condition like the partial correlation condition above, form a perfect simplex. (1954, p. 271)

Guttman's development led to a formulation of order structure among variables which did not depend on a hypothetical factor. He defined intermediacy among variables such that "a statistical variable z will be said to be intermediate to x and y if the following partial correlation vanishes: $r_{xy.z} = 0$." (Guttman, 1954, p. 273). From this he concluded that "if we are given a set of variables t_1, t_2, ... t_n, this set will be said to form a perfect simplex in this rank order if $r_{jh.k} = 0$ ($j < k < h$).

Since test scores are generally considered as comprising "true" and "error" parts, investigations into possible simplex structure of test scores should involve error considerations. Guttman has considered a number of error relationships and has stipulated corresponding types of partial simplex structures.

The investigator has adapted the above considerations to define betweeness among test variables. Patterning the definition on Guttman,

> A test t_z will be said to be <u>between</u> test t_x and t_y, if the following three conditions are true:
>
> (i) $r_{xy.z}$ is not significantly different from 0.
>
> (ii) $r_{xz.y}$ is significantly different from 0.
>
> (iii) $r_{yz.x}$ is significantly different from 0.

The relationship will be denoted (as generally in mathematics) by

$$t_x - t_y - t_z,$$

and the three variables t_x, t_y, and t_z will be said to have a

betweeness relationship among them.

Using the above definition, it is easy to see that the following theorems are true. The theorems have been patterned on the development of betweeness by Moise (1963).

Theorem 5

If $t_x - t_z - t_y$, then $t_y - t_z - t_x$.

Theorem 6

Of any three variables having a betweeness relationship among themselves, exactly one is between the other two.

The relationship may be extended to n test scores such that a betweeness relationship exists among the test scores t_1, t_2, ... t_n, in the order indicated by the subscripts if for integers $j < k < h$,

(i) $r_{jh.k}$ is not significantly different from 0,

(ii) $r_{hk.j}$ is significantly different from 0,

(iii) $r_{jk.h}$ is significantly different from 0.

When n test scores satisfy the above three conditions two rank orderings are possible, such that a test with order i in one ordering will have order (n-i) in the other ordering. Psychological considerations generally indicate the direction of the ordering. Thus the n test scores may be designated as either t_1-t_2-t_3- ... -t_n, or t_n-t_{n-1}-t_{n-2}-...-t_1.

Hierarchical Relationships

The test used to ascertain whether a particular order among test variables is a hierarchical order was Burt's (1954, p. 16) Law of increasing sign reversals, which would be satisfied by observing that the principal components of a set of variables exhibit "a progressive increase in the number of sign changes."

Burt (1954 p. 28) maintains that

> With psychological data, the appearance of a progressive sign change <u>does</u> have important psychological implications. It provides the most convincing way of verifying or refuting the hypothesis that the traits selected can be classified according to a hierarchical scheme.

<u>Betweeness and Complexity</u>

When a triple (t_1, t_2, t_3) has a betweeness relationship among them, further investigations may be made to ascertain whether the observed relationship was indicative of a complexity relationship in the Guttman sense. Such a relationship which may be signified as $t_1 \subset t_2 \subset t_3$
signifies that t_2 "requires everything t_1 does and more" and "t_3 is more complex than t_2, requiring everything t_2 does and more."

Guttman's development suggests that in this case, if the test variables are analysed for multiple prediction among themselves, it should be found that "non-neighbouring" regression weights are zero, in the perfect case, or approach zero in a partial case of a simplex. If β_{ij} represents the standard partial regression coefficient of test t_j for predicting t_i it should be expected in a perfect case that $\beta_{31} = \beta_{13} = 0$, but that every other regression weight is non-zero. This is a logical development in view of the supposition that a neighbouring test contains all the variance that a non-neighbouring test could use in predicting, and could also do the predicting with less "error."

From the above considerations and from considerations of sampling error, the following procedure was used to test the hypothesis that an observed betweeness relationship was indicative of a complexity relationship.

Procedure   A particular betweeness relationship was considered as indicative of a complexity relationship if

(i) β_{13} and β_{31} were not significantly different from zero

(ii) All other beta coefficients were significantly different from zero.

Hypothesis 5

That DP measures of production may be hierarchically ordered within each product category, in an order indicating betweeness and complexity relationships.

Analysis and Results

The hypothesis was tested in three stages. The existence of betweeness relationships among the measures in a product category was investigated in the first stage. The hypothesis was rejected for a product category, if a betweeness relationship was not found among the DP measures in that product category. In investigating betweeness relationships, a two tailed t test was used for investigating the significance of a partial correlation coefficient. The .05 level was used in all tests of significance in this study. The required t was

$$t = r_{12.3} \sqrt{\frac{N - 3}{1 - r^2_{12.3}}},$$

with N-3 degrees of freedom (Fergusson, 1966, p. 390), where N represented the number of observations, and $r_{12.3}$ represented the partial correlation coefficient of a first variable with a second where a third variable is held constant. The critical value for t in main sample 1 was $t_{.975}(37) \simeq 2.03$,

and that for main sample 2 was $t_{.975}(59) \simeq 2.00$.

The hypothesis was tested using the three DP measures (facility, variety, and novelty) in each of the product categories of units, systems, transformations, and implications. The variables used were the "stable" tests described in table XX.

The intercorrelations of the DP measures in each product category may be found in table XXI. A summary of the betweeness analysis of the DP measures within product categories may be found in table XXII. The symbols F, V, and N, are used for Facility, Variety, and Novelty measures, respectively. Partial correlation coefficients and corresponding observed t values are given under the designated columns. The decision on the significance of a partial correlation coefficient is denoted by the symbol NS or the symbol S under the D column, according as the appropriate partial correlation coefficient is not or is significantly different from zero.

Results -- Stage 1

Betweeness relationships were found among the DP measures of units, systems, and implications respectively, as may be seen from Table XXII. No betweeness relationship was found among the transformation measures of main sample 1, and the units measures of main sample 2.

Stage 2

An investigation was carried out in the second stage to find out whether the observed order of betweeness was a hierarchical order. The hypothesis was rejected if the principal components of the variables did not show a progressive increase of sign changes corresponding to the order of betweeness. The principal components of the measures are given in Table XXIII.

TABLE XXI

INTERCORRELATIONS OF DP MEASURES WITHIN PRODUCT CATEGORIES

MAIN SAMPLE 1

UNITS

	Facility	Variety	Novelty
Facility		0.82	0.69
Variety			0.75
Novelty			

SYSTEMS

	Facility	Variety	Novelty
Facility		0.90	0.94
Variety			0.82

TRANSFORMATIONS

	Facility	Variety	Novelty
Facility		0.77	0.90
Variety			0.95

IMPLICATIONS

	Facility	Variety	Novelty
Facility		0.69	0.74
Variety			0.80

MAIN SAMPLE 2

UNITS

	Facility	Variety	Novelty
Facility		0.76	0.71
Variety			0.79

TABLE XXII

SUMMARY OF BETWEENESS ANALYSIS FOR DP MEASURES WITHIN PRODUCT CATEGORY

(MAIN SAMPLE 1)

PRODUCT	$r_{FV.N}$	OBS. t	$D_{FV.N}$	$r_{VN.F}$	OBS. t	$D_{VN.F}$	$r_{FN.V}$	OBS t	$D_{VN.F}$	FINAL DECISION
Units	0.62	4.84	S	0.45	3.11	S	0.20	1.22	NS	F-V-N
Systems	0.66	5.39	S	-0.19	-1.20	NS	0.82	8.75	S	N-F-V
Trans.	-0.52	-3.67	S	0.90	12.85	S	0.79	7.91	S	REJECT
Imp.	0.24	1.51	NS	0.59	4.50	S	0.43	2.91	S	F-N-V

(MAIN SAMPLE 2)

PRODUCT	$r_{FV.N}$	OBS. t	$D_{FV.N}$	$r_{VN.F}$	OBS. t	$D_{VN.F}$	$r_{FN.V}$	OBS t	$D_{VN.F}$	FINAL DECISION
Units	0.46	3.96	S	0.54	5.00	S	0.28	2.23	S	REJECT

TABLE XXIII

PRINCIPAL COMPONENTS OF DP MEASURES-WITHIN-PRODUCTS POSSESSING BETWEENESS RELATIONSHIPS

UNITS	PRINCIPAL COMPONENTS		
	P_1	P_2	P_3
Facility	0.91	0.32	0.24
Variety	0.94	0.11	-0.33
Novelty	0.89	-0.45	0.09
Number of Sign Changes	0	1	2

SYSTEMS	PRINCIPAL COMPONENTS		
	P_1	P_2	P_3
Novelty	0.96	0.27	0.11
Facility	0.99	0.06	-0.16
Variety	0.94	-0.33	0.06
Number of Sign Changes	0	1	2

IMPLICATIONS	PRINCIPAL COMPONENTS		
	P_1	P_2	P_3
Facility	0.89	0.46	0.09
Novelty	0.93	-0.13	-0.34
Variety	0.91	-0.32	0.26
Number of Sign Changes	0	1	2

Results Stage 2

It was found that in each case in which variables possessing a betweeness relationship were analyzed into principal components, the order of sign change was 0, 1, 2, indicating a hierarchical order.

Stage 3

An investigation was conducted in the third stage to find out whether the observed order of betweeness was also an order of complexity. The hypothesis was not-rejected only if for each set of three variables considered:

(i) The two non-neighbouring standard regression coefficients were not significantly different from zero,

(ii) All other partial regression coefficients were significantly different from zero.

The following considerations were involved in the testing of the significance of a standard partial regression coefficient. For a set of m variables X_1, X_2, ..., X_m, the standard partial regression coefficient of X_j in the regression equation for predicting X_1 from the remaining (m - 1) variables, was denoted by $\beta_{1j.\underline{v}}$, where $\underline{v}$ denoted the (m - 2) subscripts (2, 3, ..., j - 1, j + 1, ..., m). A t test was used for testing that a standard partial regression coefficient was significantly different from zero. This test was based on Bartlett's development as reported by Morrison (1967, p. 105). The derived t was

$$t = \beta_{1j.\underline{v}} \sqrt{\frac{(N-m)(1 - R^2_{j.\underline{v}})}{1 - R^2_{1.j,\underline{v}}}}$$

with (N - m) degrees of freedom. Here N represented the number of

observations, and $R_{j.\underline{v}}$ represented the multiple correlation with X_j as the dependent variable, and the variables whose subscripts are represented in $\underline{v}$ as the independent variables or predictors.

The symbol β_{ij} is used here as a simple way of referring to the standard partial regression coefficient of variable X_j in the regression equation for predicting X_i from the remaining variables. The coefficients (β_{ij}) for the DP measures within each product category are given in the (a) sections of table XXIV.

The decision rule in each case was to reject the hypothesis that an observed β_{ij} was significantly different from zero, if and only if, the absolute value of the observed t value corresponding to a β_{ij} was greater than $t_{.975}(37) \cong 2.03$. The observed t values are given in the (b) sections of table XXIV. The corresponding decisions are recorded on the (c) sections of table XXIV. The symbols NS and S were used in recording the decisions to indicate that a β_{ij} corresponding to the row and column of the symbol, was not or was significantly different from zero.

Results -- Stage 3

It was observed that in each case,

(i) the two non-neighbouring regression coefficients were not significantly different from zero,

(ii) all other regression coefficients were significantly different from zero.

It was concluded that in each case the observed order of betweeness was an order of complexity.

Direction of Complexity

The analysis on complexity did not indicate the direction

TABLE XXIV

SUMMARY OF COMPLEXITY ANALYSIS FOR DP MEASURES WITHIN PRODUCT CATEGORIES

(i) UNITS

(a) STANDARD PARTIAL REGRESSION COEFFICIENTS (β_{ij})

	F	V	N	Multiple R	Max r
F	--	0.69	0.17	0.82	0.82
V	0.57	--	0.36	0.86	0.82
N	0.22	0.57	--	0.76	0.75

(b) OBSERVED t VALUES CORRESPONDING TO β_{ij}

	F	V	N
F	--	4.84	1.22
V	4.84	--	3.11
N	1.22	3.11	--

(c) DECISIONS ON SIGNIFICANCE OF β_{ij} $t_{.975}(37) \simeq 2.03$

	F	V	N
F	-	S	NS
V	S	-	S
N	NS	S	-

TABLE XXIV (continued)

(ii) SYSTEMS

(a) STANDARD PARTIAL REGRESSION COEFFICIENTS (β_{ij})

	N	F	V	Multiple R	Max r
N	--	1.08	-0.15	0.95	0.94
F	0.63	--	0.39	0.97	0.94
V	-0.25	1.14	--	0.90	0.90

(b) OBSERVED t VALUES CORRESPONDING TO β_{ij}

	N	F	V
N	--	8.75	-1.2
F	8.75	--	5.39
V	-1.20	5.39	--

(c) DECISIONS ON SIGNIFICANCE OF β_{ij} $t_{.975}(37) \simeq 2.03$

	N	F	V
N	-	S	NS
F	S	-	S
V	NS	S	-

TABLE XXIV (continued)

(iii) IMPLICATIONS

(a) STANDARD PARTIAL REGRESSION COEFFICIENTS (β_{ij})

	N	F	V	Multiple R	Max r
N	--	0.52	0.27	0.76	0.74
F	0.36	--	0.55	0.84	0.80
V	0.21	0.64	--	0.81	0.80

(b) OBSERVED t VALUES CORRESPONDING TO β_{ij}

	F	N	V
F	--	2.92	1.51
N	2.92	--	4.50
V	1.51	4.50	--

(c) DECISIONS OF SIGNIFICANCE OF β_{ij} $t_{.975}(37) \simeq 2.03$

	F	N	V
F	-	S	NS
N	S	-	S
V	NS	S	-

of increasing complexity. Direction was determined on non-statistical grounds. It was assumed that variety was more complex than facility, while facility was measured as the number of appropriate responses made by a student to a test situation, variety was measured by taking all the appropriate responses to a test situation into account, and determining a reduced set of ideas that accounted for the observed ideas.

Results and Discussion

With the above assumption, the results of the betweeness and complexity analyses were as follows:

PRODUCT	BETWEENESS	COMPLEXITY
UNITS	F-V-N	$F \subset V \subset N$
SYSTEMS	N-F-V	$N \subset F \subset V$
IMPLICATIONS	F-N-V	$F \subset N \subset V$

No uniform order was found among the measures across products. One explanation of the observed orders may be based on the productivity levels for the composite tests which made up the facility measures in each product category. When productivity level for a composite test was calculated as the proportion of the sample whose average score was greater than one, it was found that the productivity levels for the sample under consideration (main sample 1) were 1.00, 0.43, and 0.73 for units, systems, and implications respectively. Thus a possible interpretation could be based on observing that where productivity was high, as in units, novelty was the most complex quality; where productivity was low, as in systems, novelty was essential in every quality, and hence was the most basic quality, and variety, involving novelty, was the most complex quality. When productivity

was moderate, as in implications, novelty, though not the most basic quality, was still a pre-requisite for variety.

Hypothesis 6

That product categories may be hierarchically ordered within each DP measure.

Analysis and Results -- Stage 1

The hypothesis was tested in three stages. In stage 1, betweeness relations were investigated among the products within each of the measures of facility, variety, and novelty. The analyses were conducted using the "stable" tests of main sample 1.

The analysis for identifying betweeness relationships was conducted using all the possible selections of three variables for each measure. Using the following symbols for units (U), classes (C), systems (S), transformations (T), and implications (I), there were four facility variables, (U,S,T, and I), five variety variables (U,C,S,T, and I), and five novelty variables (U,C,S,T, and I). Accordingly there were four, ten, and ten possible selections of three facility, variety, and novelty variables respectively.

Each selection of three was tested for a betweeness relationship, in each case using a two tailed t test for the significance of a partial correlation coefficient at the .05 level, as explained in the analysis of hypothesis 5.

The intercorrelations of the products within each DP measure are given in table XXV. A summary of the betweeness analysis is given in table XXVI.

TABLE XXV

INTERCORRELATIONS OF PRODUCTS WITHIN DP MEASURES - MAIN SAMPLE 1 - "STABLE" TESTS

(a) FACILITY MEASURE

	UNITS	SYSTEMS	TRANS.	IMP.
UNITS		0.51	0.42	0.44
SYSTEMS			0.73	0.33
TRANS.				0.41
IMP.				

(b) VARIETY MEASURE

	UNITS	CLASSES	SYSTEMS	TRANS.	IMP.
UNITS		0.29	0.35	0.13	0.18
CLASSES			0.44	0.24	0.16
SYSTEMS				0.82	0.65
TRANS.					0.64
IMP.					

(c) NOVELTY MEASURE

	UNITS	CLASSES	SYSTEMS	TRANS.	IMP.
UNITS		0.42	0.40	0.23	0.26
CLASSES			0.50	0.38	0.30
SYSTEMS				0.75	0.26
TRANS.					0.33
IMP.					

TABLE XXVI

SUMMARY OF BETWEENESS ANALYSIS FOR
PRODUCTS WITHIN DP MEASURES

FACILITY MEASURE

S	X_1	X_2	X_3	$r_{12.3}$	t (obs)	$D_{12.3}$	$r_{23.1}$	t (obs)	$D_{23.1}$	$r_{31.2}$	t (obs)	$D_{31.2}$	Final Decision
1	U	S	T	0.33	2.09	S	0.66	5.30	S	0.09	0.53	NS	U-S-T
2	U	S	I	0.43	2.89	S	0.14	0.84	NS	0.33	2.14	S	I-U-S
3	U	T	I	0.30	1.89	NS	0.27	1.68	NS	0.32	2.07	S	REJECT
4	S	T	I	0.69	5.76	S	0.25	1.58	NS	0.06	0.35	NS	REJECT

VARIETY MEASURE

S	X_1	X_2	X_3	$r_{12.3}$	t (obs)	$D_{12.3}$	$r_{23.1}$	t (obs)	$D_{23.1}$	$r_{31.2}$	t (obs)	$D_{31.2}$	Final Decision
1	U	C	S	0.16	1.00	NS	0.38	2.50	S	0.25	1.60	NS	REJECT
2	U	C	T	0.27	1.68	NS	0.22	1.36	NS	0.06	0.39	NS	REJECT
3	U	C	I	0.27	1.69	NS	0.11	0.69	NS	0.14	0.87	NS	REJECT
4	U	S	T	0.42	2.84	S	0.84	9.25	S	-0.29	-1.83	NS	U-S-T
5	U	S	I	0.31	1.95	NS	0.64	5.04	S	-0.06	-0.39	NS	REJECT
6	U	T	I	0.20	0.12	NS	0.63	4.98	S	0.13	0.77	NS	REJECT
7	C	S	T	0.44	2.94	S	0.82	8.76	S	-0.23	-1.45	NS	C-S-T
8	C	S	I	0.45	3.08	S	0.66	5.29	S	-1.19	-1.18	NS	C-S-I
9	C	T	I	0.19	1.17	NS	0.63	4.93	S	0.00	0.01	NS	REJECT
10	S	T	I	0.70	5.88	S	0.24	1.53	NS	0.28	1.80	NS	REJECT

NOVELTY MEASURE

S	X_1	X_2	X_3	$r_{12.3}$	t (obs)	$D_{12.3}$	$r_{23.1}$	t (obs)	$D_{23.1}$	$r_{31.2}$	t (obs)	$D_{31.2}$	Final Decision
1	U	C	S	0.27	1.73	NS	0.40	2.63	S	0.24	1.52	NS	REJECT
2	U	C	T	0.36	2.37	S	0.32	2.08	S	0.09	0.99	NS	U-C-T
3	U	C	I	0.37	2.40	S	0.22	1.34	NS	0.15	0.95	NS	REJECT
4	U	S	T	0.35	2.26	S	0.74	6.61	S	-0.11	-0.67	NS	U-S-T
5	U	T	I	0.16	1.00	NS	0.28	1.79	NS	0.20	1.23	NS	REJECT
6	U	S	I	0.35	2.31	S	0.18	1.10	NS	0.17	1.07	NS	REJECT
7	C	S	T	0.34	2.22	S	0.70	5.91	S	0.02	0.11	NS	C-S-T
8	C	S	I	0.45	3.10	S	0.14	0.84	NS	0.20	1.24	NS	REJECT
9	C	T	I	0.32	2.03	NS	0.24	1.50	NS	0.20	1.22	NS	REJECT
10	S	T	I	0.73	6.44	S	0.20	1.26	NS	0.03	0.17	NS	REJECT

The results of the betweeness analysis were as follows:

MEASURE	BETWEENESS
Facility	U-S-T
	I-U-S
Variety	U-S-T
	C-S-T
	C-S-I
Novelty	U-S-T
	U-C-T
	C-S-T

Stage 2

The second stage of the analysis was to verify that the order of betweeness relationships found was hierarchical. The principal components of the product variables among which betweeness relationships were found, are given in table XXVII. It was found that in each case, the order of sign change was 0, 1, 2, indicating a hierarchical order in each case.

Stage 3

The third stage was to investigate that the order of betweeness was also an order of complexity. The hypothesis was tested under the same conditions as outlined in stage 3 of the analysis of hypothesis 5.

The standard partial regression coefficients of the products within each of the three measures of facility, variety, and novelty,

TABLE XXVII

PRINCIPAL COMPONENTS OF DP PRODUCTS-WITHIN-MEASURES POSSESSING BETWEENESS RELATIONSHIPS

FACILITY

	(i) PRINCIPAL COMPONENTS				(ii) PRINCIPAL COMPONENTS		
	P_1	P_2	P_3		P_1	P_2	P_3
Units	0.74	0.67	0.07	Imp.	0.73	0.65	0.20
Systems	0.90	-0.20	-0.38	Units	0.84	-0.11	-0.53
Trans.	0.87	-0.37	0.34	Systems	0.78	-0.49	0.38
Number of Sign Changes	0	1	2		0	1	2

VARIETY

	(i) PRINCIPAL COMPONENTS				(ii) PRINCIPAL COMPONENTS		
	P_1	P_2	P_3		P_1	P_2	P_3
Units	0.47	0.88	0.07	Classes	0.60	0.79	0.07
Systems	0.96	-0.10	-0.28	Systems	0.95	-0.14	-0.28
Trans.	0.90	-0.36	0.26	Trans.	0.88	-0.39	0.25
Number of Sign Changes	0	1	2		0	1	2

	(iii) PRINCIPAL COMPONENTS		
	P_1	P_2	P_3
Classes	0.61	0.77	0.17
Systems	0.92	-0.07	-0.39
Imp.	0.80	-0.51	0.32
Number of Sign Changes	0	1	2

TABLE XXVII (continued)

NOVELTY MEASURE

(i)

	PRINCIPAL COMPONENTS		
	P_1	P_2	P_3
Units	0.73	0.58	0.37
Classes	0.82	0.04	-0.57
Trans.	0.70	-0.66	0.29
Number of Sign Changes	0	1	2

(ii)

	PRINCIPAL COMPONENTS		
	P_1	P_2	P_3
Classes	0.71	0.70	0.08
Systems	0.91	-0.18	-0.36
Trans.	0.87	-0.38	0.32
Number of Sign Changes	0	1	2

(iii)

	PRINCIPAL COMPONENTS		
	P_1	P_2	P_3
Units	0.59	0.80	0.09
Systems	0.92	-0.15	-0.35
Trans.	0.87	-0.39	0.31
Number of Sign Changes	0	1	2

are given in the (a) sections of tables XXVIII, XXIX, and XXX. The observed t values corresponding to β_{ij}, are given in the (b) sections of the tables. The decisions on the significance of each standard partial regression coefficient using a two-tailed t test at the .05 level of significance are given in the (c) sections of tables XXVIII, XXIX, and XXX.

It was found that in each case where a betweeness relationship was observed,

(i) the two non-neighbouring regression coefficients were not significantly different from zero.

(ii) all other regression coefficients were significantly different from zero.

It was concluded that in each case the observed order of betweeness was an order of complexity.

Direction of Complexity

The direction of increasing complexity was determined by using the "Guilford order" of units, classes, relations, systems, transformations, and implications, with units being of lowest rank, as a working hypothesis.

Results

The results of the betweeness and complexity analyses were as follows:

MEASURE	BETWEENESS	COMPLEXITY
Facility	U-S-T	$U \subset S \subset T$
	I-U-S	$I \subset U \subset S$
Variety	U-S-T	$U \subset S \subset T$
	C-S-T	$C \subset S \subset T$
	C-S-I	$C \subset S \subset I$

TABLE XXVIII

SUMMARY OF COMPLEXITY ANALYSIS FOR PRODUCTS WITHIN FACILITY MEASURES

1(a) STANDARD PARTIAL REGRESSION COEFFICIENTS (β_{ij})

	Units	Systems	Trans.	Multiple R	Max r
U	--	0.43	0.11	0.51	0.51
S	0.25	--	0.62	0.76	0.73
T	0.07	0.69	--	0.73	0.73

1(b) OBSERVED t VALUES CORRESPONDING TO β_{ij}

	Units	Systems	Trans.
U	--	2.09	0.53
S	2.09	--	5.30
T	0.53	5.30	--

1(c) DECISIONS ON SIGNIFICANCE OF β_{ij} $t_{.975}(37) \approx 2.03$

	Units	Systems	Trans.
U	--	S	NS
S	S	--	S
T	NS	S	--

TABLE XXVIII (continued)

2(a) STANDARD PARTIAL REGRESSION COEFFICIENTS (β_{ij})

	Imp.	Units	Systems	Multiple R	Max r
I	--	0.36	0.14	0.45	0.44
U	0.30	--	0.41	0.58	0.51
S	0.13	0.45	--	0.52	0.51

2(b) OBSERVED t VALUES CORRESPONDING TO (β_{ij})

	Imp.	Units	Systems
I	--	2.14	0.84
U	2.14	--	2.89
S	0.84	2.89	--

2(c) DECISIONS ON SIGNIFICANCE OF β_{ij} $t_{.975}(37) \simeq 2.03$

	Imp.	Units	Systems
I	--	S	NS
U	S	--	S
S	NS	S	--

TABLE XXIX

SUMMARY OF COMPLEXITY ANALYSIS FOR PRODUCTS WITH VARIETY MEASURES

1(a) STANDARD PARTIAL REGRESSION COEFFICIENTS (β_{ij})

	Units	Systems	Trans.	Multiple R	Max r
U	--	0.74	-0.48	0.44	0.35
S	0.24	--	0.79	0.86	0.82
T	-0.17	0.88	--	0.84	0.82

1(b) OBSERVED t VALUES CORRESPONDING TO (β_{ij})

	Units	Systems	Trans.
U	--	2.84	-1.83
S	2.84	--	9.26
T	-1.83	9.26	--

1(c) DECISIONS ON SIGNIFICANCE OF (β_{ij}) CRITICAL VALUE: $t_{.975}(37) \cong 2.03$

	Units	Systems	Trans.
U	--	S	NS
S	S	--	S
T	NS	S	NS

TABLE XXIX (continued)

2(a) STANDARD PARTIAL REGRESSION COEFFICIENTS (β_{ij})

	Classes	Systems	Trans.	Multiple R	Max r
C	--	0.74	-0.37	0.49	0.44
S	0.26	--	0.76	0.86	0.83
T	-0.15	0.89	--	0.83	0.82

2(b) OBSERVED t VALUES CORRESPONDING TO (β_{ij})

	Classes	Systems	Trans.
C	--	2.94	-1.45
S	2.94	--	8.76
T	-1.45	8.76	--

2(c) DECISIONS OF SIGNIFICANCE OF (β_{ij})

(CRITICAL VALUE: $t_{.975}(37) \simeq 2.03$)

	Classes	Systems	Trans.
C	--	S	NS
S	S	--	S
T	NS	S	--

TABLE XXIX (continued)

3(a) STANDARD PARTIAL REGRESSION COEFFICIENTS (β_{ij})

	Classes	Systems	Implications	Multiple R	Max r
C	--	0.59	-0.23	0.47	0.44
S	0.35	--	0.60	0.74	0.65
I	-0.16	0.72	--	0.67	0.65

3(b) OBSERVED t VALUES CORRESPONDENCE TO (β_{ij})

	Classes	Systems	Implications
C	--	3.08	-1.18
S	3.08	--	5.29
I	-1.18	5.29	--

3(c) DECISIONS ON SIGNIFICANCE OF (β_{ij})

(CRITICAL VALUE: $t_{.975}(37) \simeq 2.03$)

	Classes	Systems	Implications
C	--	S	NS
S	S	--	S
I	NS	S	NS

TABLE XXX

SUMMARY OF COMPLEXITY ANALYSIS FOR PRODUCTS WITHIN NOVELTY MEASURES

1(a) STANDARD PARTIAL REGRESSION COEFFICIENTS (β_{ij})

	Units	Systems	Trans.	Multiple R	Max r
U	--	0.51	-0.15	0.41	0.40
S	0.24	--	0.69	0.78	0.75
T	-0.08	0.78	--	0.75	0.75

1(b) OBSERVED t VALUES CORRESPONDING TO (β_{ij})

	Units	Systems	Trans.
U	--	2.26	-0.67
S	2.26	--	6.61
T	-0.67	6.61	--

1(c) DECISIONS ON SIGNIFICANCE OF (β_{ij})

(CRITICAL VALUE: $t_{.975}(37) \simeq 2.03$)

	Units	Systems	Trans.
U	--	S	NS
S	S	--	S
T	NS	S	--

TABLE XXX (continued)

2(a) STANDARD PARTIAL REGRESSION COEFFICIENTS (β_{ij})

	Units	Classes	Trans.	Multiple R	Max r
U	--	0.38	0.09	0.42	0.42
C	0.35	--	0.30	0.51	0.42
T	0.09	0.35	--	0.39	0.38

2(b) OBSERVED t VALUES CORRESPONDING TO (β_{ij})

	Units	Classes	Trans.
U	--	2.94	-1.45
C	2.94	--	8.76
T	-1.45	8.76	--

2(c) DECISIONS ON SIGNIFICANCE OF (β_{ij})

(CRITICAL VALUE: $t_{.975}(37) \doteq 2.03$)

	Units	Classes	Trans.
U	--	S	NS
C	S	--	S
T	NS	S	--

TABLE XXX (continued)

3(a) STANDARD PARTIAL REGRESSION COEFFICIENTS (β_{ij})

	Classes	Systems	Trans.	Multiple R	Max r
C	--	0.48	0.02	0.50	0.50
S	0.25	--	0.65	0.78	0.75
T	0.01	0.74	--	0.75	0.75

3(b) OBSERVED t VALUES CORRESPONDING TO (β_{ij})

	Classes	Systems	Trans.
C	--	2.22	0.22
S	2.22	--	5.91
T	0.11	5.91	--

3(c) DECISIONS ON SIGNIFICANCE OF (β_{ij})

(CRITICAL VALUE: $t_{.975}(37) \simeq 2.03$)

	Classes	Systems	Trans.
C	--	S	NS
S	S	--	S
T	NS	S	--

MEASURE	BETWEENESS	COMPLEXITY
Novelty	U-S-T	$U \subset S \subset T$
	U-C-T	$U \subset C \subset T$
	C-S-T	$C \subset S \subset T$

It was found that the Guilford order was tenable for all but one of the triples of product categories for which betweeness--complexity relationships were demonstrated. The one exception was the facility I-U-S relationship. Since the Guilford order puts units as basic, units could not then occupy a rank between any other two products. However Guilford has suggested that there could be some change in the original ranking of implications: "There might be some sense in putting implications immediately units, since implications are the simplest and most general way in which units can be connected." (Guilford, 1967, p. 63.) It is to be noted, however, that the complexity order adopted here of $I \subset U \subset S$, does contradict the Guilford order. This order was adopted because the alternative of $S \subset U \subset I$, was considered even less plausible, since systems was conceived as involving "the ability to organize elementary ideas into complex ones" (See section 3.3).

Further Development

The number of betweeness triples found for each DP measure was too few to indicate a betweeness relationship among more than three variables. The smallest number of variables needed to establish a betweeness relation among n variables when tested three at a time is ${}^{n}C_{3}$. Thus, at least four betweeness triples were necessary to establish a betweeness relation among four products in each measure.

An examination of the complexity relationships for each measure indicated certain complexity relationships among more than three products which would have been true relationships if the original complexity relationships which were used to build them up had been without error. A combination of the facility relationships found resulted in the projection: $I \subseteq U \subseteq S \subseteq T$. There was insufficient evidence to project a complexity relationship among more than three variety products. A combination of the observed novelty relationships resulted in the projection: $U \subseteq C \subseteq S \subseteq T$.

Since each of the original complexity relationships from which the projected complexity quadruples were built up were subject to error, there was no compulsion about the projections, and they were tested for hierarchical and complexity relationships. The principal components for the two sets of variables may be found in table XXXI. It was found that the sign changes for the facility principal components increased progressively -- 0, 1, 2, 3, and indicated a hierarchical order. The principal components of the novelty variables and a non-decreasing progression in the number of sign changes -- 0, 1, 2, 2, and this indicated some, although not very strong evidence of a hierarchy.

A summary of the complexity analysis for the facility and novelty quadruples may be found in table XXXII. It was found that in the case of the facility variables, the non-neighbouring regression coefficients were non significant, while the neighbouring regression coefficients were significant, indicating a complexity relationship. The analysis for the novelty quadruple did not indicate a complexity relationship

TABLE XXXI

PRINCIPAL COMPONENTS OF PROJECTED COMPLEXITY QUADRUPLES

FACILITY

	PRINCIPAL COMPONENTS			
	P_1	P_2	P_3	P_4
Imp.	0.67	0.65	0.36	0.09
Units	0.75	0.25	-0.60	-0.10
Systems	0.85	-0.40	-0.02	0.35
Trans.	0.84	-0.33	0.27	-0.33
Number of Sign Changes	0	1	2	3

NOVELTY

	P_1	P_2	P_3	P_4
Units	0.62	0.67	0.40	0.07
Classes	0.74	0.30	-0.61	0.04
Systems	0.89	-0.26	0.11	-0.36
Trans.	0.80	-0.50	0.12	0.30
Number of Sign Changes	0	1	2	2

TABLE XXXII

SUMMARY OF COMPLEXITY ANALYSIS OF PROJECTED COMPLEXITY QUADRUPLES

FACILITY MEASURES

(a) STANDARD PARTIAL REGRESSION COEFFICIENTS (β_{ij})

	Imp.	Units	Systems	Trans.	Multiple R
I	--	0.34	-0.09	0.33	0.51
U	0.30	--	0.41	0.00	0.58
S	-0.05	0.26	--	0.64	0.80
T	0.20	0.00	0.66	--	0.75

(b) OBSERVED t VALUES CORRESPONDING TO (β_{ij})

	Imp.	Units	Systems	Trans.
I	--	2.04	-0.40	1.59
U	2.04	--	2.09	-0.02
S	-0.40	2.09	--	5.14
T	1.59	-0.02	5.14	--

(c) DECISIONS ON SIGNIFICANCE OF (β_{ij}) CRITICAL VALUE: $t_{.975}(37) \simeq 2.03$

	Imp.	Units	Systems	Trans.
I	--	S	NS	NS
U	S	--	S	NS
S	NS	S	--	S
T	NS	NS	S	--

TABLE XXXII (continued)

NOVELTY MEASURE

(a) STANDARD PARTIAL REGRESSION COEFFICIENTS (β_{ij})

	Units	Classes	Systems	Trans.	Multiple R
U	--	0.29	0.37	-0.16	0.48
C	0.26	--	0.34	0.06	0.55
S	0.18	0.18	--	0.64	0.80
T	-0.09	0.04	0.76	--	0.75

(b) OBSERVED t VALUES CORRESPONDING TO (β_{ij})

	Units	Classes	Systems	Trans.
U	--	1.73	1.59	-0.72
C	1.73	--	1.53	0.30
S	1.59	1.53	--	5.87
T	-0.72	0.30	5.87	--

(c) DECISIONS ON SIGNIFICANCE OF (β_{ij}) CRITICAL VALUE: $t_{.975}(37) \simeq 2.03$

	Units	Classes	Systems	Trans.
U	--	NS	NS	NS
C	NS	--	NS	NS
S	NS	NS	--	S
T	NS	NS	S	--

6.3-24 Discussion -- Hierarchical Relationships

Analyses have been conducted to reveal possible hierarchies among variables, indicating betweeness and complexity relationships. Since hierarchical relationships can be shown to exist among variables in many ways, it was considered essential that if the existence of such a relationship should be used as evidence of psychological structure, the hypothesized structural relationships of the variables should be specified in such a way that the number of different ways in which they could be evident among a set of variables would be small.

A betweeness relationship has been defined among variables, in a way which follows Guttman's (1954) notion of intermediacy among variables. The betweeness relationship was such that if it existed among variables, the variables could be arranged in two ways indicating the relationship, such that the order of the variables in one arrangement is the exact opposite of the order in the other arrangement. The complexity relationship has been used to investigate whether qualities which variables measure were subsumed in other qualities in a betweeness continuum.

The measures of facility, variety, and novelty were found to have a betweeness-complexity hierarchical relationship with regard to the products, units, systems, and implications. No betweeness relationship was found among the DP measures of the transformation category. There was no uniform order found among the measures across the products. An explanation suggested was an interpretation of the findings in terms of the productivity levels of the measures. Where the students found it easy to produce ideas, novelty was the

most complex quality, but where production was difficult, variety was the most complex quality.

The analysis for hierarchical relationships among the products did not reveal that all the products investigated were connected in betweeness-complexity hierarchies within the DP measures. However, two betweeness-complexity triples were found among the four facility variables, three were found among the five variety variables, and three were found among four of the five novelty variables. The Guilford (1967) "logical order" of units, classes, systems, relations, transformations, and implications, was tenable in all but one of these triples. The exception occurred with facility in implications. The analysis indicated that this quality was of a lower order than facility in units. Guilford himself seems to have questioned the position of implications, and suggested an alternative placing next to units, with units as the most basic. It is interesting to note that variety in implications was found to be a rather complex ability among the variety measures.

The findings that most of the hierarchical relationships observed did not contradict Guilford's "logical order" is of importance in some ways to the construct validity of the "stable" tests. The findings indicate that the underlying constructs may be related in a way that is logically plausible.

6.3-25 Summary and Conclusion - Construct Validation

The tests constructed by the investigator have been examined experimentally to find out whether their underlying structure corresponded to certain logical expectations. Confirmation of such correspondence was considered to be evidence of construct validity.

These expectations have included factorial groupings, prediction of subject matter, and hierarchical orderings in terms of increasing complexity. It has been found that the tests met these expectations to a large extent, and that the expectations were more clearly met among the sample of volunteers than among the sample of regular students.

6.4 CRITERION VALIDATION

Validation procedures which involved the criterion of inventiveness of the study have been classified as criterion validation procedures. The criterion for validation used in this study was the mean rating of the SIGNIFICANT GROUP. (See sections 3.5-4, and 4.4-1).

6.4-1 Ratings of the SIGNIFICANT GROUP

The SIGNIFICANT GROUP gave ratings to the response categories of the fourteen DP tests administered. A response category was a student's response as interpreted and standardized by the investigator. The reliability estimate for the average rating of the response categories of each test, was obtained using analysis of variance techniques. (Winer, 1962). Each response category was treated as an "individual." The variation due to the judges was eliminated from the within-response variation, and the residual variation so obtained was treated as error variation. Summaries of analysis of variance for each test and the obtained reliability estimates may be found in table XXXIII. It was found that the reliability estimate in each case was significantly different from zero, and the observed reliability estimates ranged from 0.64 to 0.89.

TABLE XXXIII

SUMMARY OF ANALYSIS OF VARIANCE FOR OBTAINING RELIABILITY ESTIMATE FOR AVERAGE RATING OF RESPONSE CATEGORIES BY SIGNIFICANT GROUP FOR EACH DP TEST

(i) Test I. Reliability estimate: 0.87

Summary of Analysis of Variance

Source of Variation	SS	df	MS	F	P
Between Response Categories	216.59	86	2.52	7.99	<.05
Residual (Error)	187.50	595	0.32		

(ii) Test 1. Reliability estimate: 0.88

Summary of Analysis of Variance

Source of Variation	SS	df	MS	F	P
Between Response Categories	326.35	92	3.55	8.09	<.05
Residual (Error)	277.92	634	0.44		

(iii) Test IB. Reliability estimate: 0.79

Summary of Analysis of Variance

Source of Variation	SS	df	MS	F	P
Between Response Categories	49.27	32	1.54	4.73	<.05
Residual (Error)	72.29	222	o.33		

TABLE XXXIII (continued)

(iv) Test IB. Reliability estimate: 0.84

Summary of Analysis of Variance

Source of Variation	SS	df	MS	F	P
Between Response Categories	16.79	13	1.29	6.43	$<.05$
Residual (Error)	18.28	91	0.20		

(v) Test II. Reliability estimate: 0.88

Summary of Analysis of Variance

Source of Variation	SS	df	MS	F	P
Between Response Categories	9.84	3	3.28	8.20	$<.05$
Residual (Error)	8.41	21	0.40		

(vi) Test 2. Reliability estimate: 0.86

Summary of Analysis of Variance

Source of Variation	SS	df	MS	F	P
Between Response Categories	11.25	3	3.75	7.00	$<.05$
Residual (Error)	11.25	21	0.54		

TABLE XXXIII (continued)

(vii) Test IIB. Reliability estimate: 0.78

Summary of Analysis of Variance

Source of Variation	SS	df	MS	F	P
Between Response Categories	6.9	4	1.73	4.51	$<.05$
Residual (Error)	10.7	28	0.38		

(viii) Test IV. Reliability estimate: 0.85

Summary of Analysis of Variance

Source of Variation	SS	df	MS	F	P
Between Response Categories	17.91	5	3.58	6.69	$<.05$
Residual (Error)	18.75	35	0.54		

(ix) Test 4. Reliability estimate: 0.64

Summary of Analysis of Variance

Source of Variation	SS	df	MS	F	P
Between Response Categories	27.65	23	1.20	2.79	$<.05$
Residual (Error)	67.33	156	0.43		

TABLE XXXIII (continued)

(x) Test 4B. Reliability estimate: 0.80

Summary of Analysis of Variance

Source of Variation	SS	df	MS	F	P
Between Response Categories	123.62	63	1.96	4.93	$<.05$
Residual (Error)	158.78	399	0.40		

(xi) Test 5. Reliability estimate: 0.84

Summary of Analysis of Variance

Source of Variation	SS	df	MS	F	P
Between Response Categories	43.38	10	4.34	6.17	$<.05$
Residual (Error)	55.50	79	0.70		

(xii) Test 5B. Reliability estimate: 0.89

Summary of Analysis of Variance

Source of Variation	SS	df	MS	F	P
Between Response Categories	259.94	93	2.80	9.48	$<.05$
Residual (Error)	207.54	704	0.29		

TABLE XXXIII (continued)

(xiii) Test VI. Reliability estimate: 0.80

Summary of Analysis of Variance

Source of Variation	SS	df	MS	F	P
Between Response Categories	96.73	44	2.20	5.11	$<.05$
Residual (Error)	145.88	339	0.43		

(xiv) Test 6B. Reliability estimate: 0.80

Summary of Analysis of Variance

Source of Variation	SS	df	MS	F	P
Between Response Categories	123.62	63	1.96	4.93	$<.05$
Residual (Error)	158.78	399	0.40		

6.4-2 Criteria for Inventive Responses

The SIGNIFICANT GROUP was requested to rate non-inventive responses as zero, and inventive responses as 1, 2, 3, or 4, according to the quality of inventiveness, with 4 denoting the highest order of inventiveness. For the purpose of using the responses in decision making, the investigator classified them as follows on the basis of the average rating given by the SIGNIFICANT GROUP.

Class 35: Response categories receiving average rating r where $3.5 \leq r \leq 4$, and at least seven judges rating the response as inventive.

Class 30: Response categories receiving an average rating r where $3.0 \leq r < 3.5$, and at least six judges rating the response as inventive.

Class 25: Response categories receiving an average rating r where $2.5 \leq r < 3.0$, and at least five judges rating the response as inventive.

Class 20: Response categories receiving average rating r where $2.0 \leq r < 2.5$, and at least four judges rating the response as inventive.

Class 15: Response categories receiving average rating r where $1.5 \leq r < 2.0$, and at least three judges rating the response as inventive.

Class 10: Response categories receiving average rating r where $1.0 \leq r < 1.5$, and at least two judges rating the response as inventive.

Class 00: All other response categories.

Class 00 response categories were considered as non inventive,

and all other classes were considered as classes of inventive responses. Since most of the response categories were judged by eight judges, it worked out that class 00 responses were those responses which rated below 1 in average.

6.4-3 DP Tests and Inventive Behavior

Hypotheses 7 and 8 were tested as investigations into the effectiveness of the DP tests in evoking inventiveness in students.

Hypothesis 7

That each subject will produce at least one inventive response.

6.4-4 Preliminary Discussion - Hypothesis 7

The hypothesis was tested using the "stable" tests common to the two samples. These were tests 1 and I in the units product category, and tests 2 and II in the classes product category. A student's response was considered as inventive, if it fell in a response category such that it received an average rating of at least one from the SIGNIFICANT GROUP, where at least two judges rated the response as indicating inventiveness.

6.4-5 Analysis and Results - Hypothesis 7

A summary of the number of inventive responses produced by the students in various inventiveness classes may be found in tables 2 and 3 of Appendix A. It was found that each of the forty subjects of main sample 1 and 59 of the 62 subjects of main sample 2 produced inventive responses in the composite of tests 1 and I. The three main sample 2 students who did not produce inventive responses in tests 1 or I, were found to have produced inventive responses when the composite of the classes product category tests 2 and II were

considered. Thus every subject was found to have produced at least one inventive response, substantiating the hypothesis.

Hypothesis 8

That each DP test will evoke inventive responses.

6.4-6 Preliminary Discussion

The hypothesis was tested for all fourteen DP tests administered to at least one of the two main samples. The classification of responses reported in section 6.4-2 above was used in selecting inventive responses.

6.4-7 Analysis and Results -- Hypothesis 8

The hypothesis was considered tenable for a test if at least one inventive response was produced for the test. A summary of the number of standardized inventive responses produced in each of the inventive classes defined above may be found in table XXIV. It was found that each test evoked inventive responses.

6.4-8 Discussion -- Hypotheses 7 and 8

The above investigations indicated that students were capable of producing inventive responses, and that the methods of divergent production led to the production of some inventiveness in many students. Although some students did not produce inventive responses in some tests, all the students produced at least one inventive response in some test.

The findings that each test evoked inventiveness are of maximum importance in establishing that the tests were tests of inventiveness. The results support the proposition that divergent production leads to inventiveness.

TABLE XXXIV

NUMBERS OF INVENTIVE RESPONSE CATEGORIES PRODUCED FOR EACH TEST

Test No.	Class 00	Class 10	Class 15	Class 20	Class 25	Class 30	Class 35
I	29	25	22	9	2	0	0
1	24	26	23	9	9	2	0
IB	3	9	16	5	0	0	0
1B	6	4	4	4	0	0	0
II	0	1	0	2	1	0	0
2	0	1	0	1	2	0	0
IIB	1	0	2	2	0	0	0
IV	0	1	2	1	2	0	0
4	0	1	5	14	3	1	0
4B	0	14	20	21	7	2	0
5	1	2	0	5	2	1	0
5B	11	41	22	11	7	2	0
VI	0	10	15	14	5	1	0
6B	3	7	7	6	0	2	0

It was not all the responses of the students that were considered as inventive. Responses which fell in class 00 as defined above, were considered as non-inventive. Divergent production was not equated with inventiveness. The psychological expectation was that someone who gave himself to divergent production in mathematics would tire of trivial production and either give up or invent.

6.4-9 DP Measures and the Inventiveness Criterion

Hypotheses 9 and 10 were tested as investigations into statistical relationships between DP measures and criterion measures of inventiveness

Hypothesis 9. That each DP measure correlates significantly with the criterion measure of inventiveness.

6.4-10 Preliminary Discussion -- Hypothesis 9

The criterion measure of the inventiveness of a student in a test was obtained indirectly. The mean rating of the SIGNIFICANT GROUP of a standardized response was taken as the criterion measure of a student's response which the standardized response represented. The sum of the criterion measures of all responses made by a student for a test was taken as the criterion measure of the student's inventiveness in that test.

The testing of the hypothesis was restricted to four tests. These were Units tests 1 and I, and Classes tests II and 2. These were the four "stable" tests common to the two samples.

6.4-10 Analysis and Results -- Hypothesis 9.

The decision rule was to reject the hypothesis for each measure of a test if its simple correlation with the criterion measure was not - significantly different from zero.

The correlations of the facility, variety, and novelty measures of test 1 with the criterion measure of test 1, were found to be 0.93, 0.66, and 0.66, respectively for main sample 1, and 0.94, 0.66, and 0.76, respectively for main sample 2. Each of these correlation coefficients are significant at the .05 level, where 38 degrees of freedom was allowed for main sample 1, and 60 degrees of freedom was allowed for main sample 2. The hypothesis was considered supported for test 1.

The correlations of the facility, variety, and novelty measures of test I with the criterion measure of test I, were found to be 0.94, 0.74, and 0.64 respectively for main sample 1, and 0.94, 0.84, and 0.67 respectively, for main sample 2. Each of these correlation coefficients was found to be significant at the .05 level, where 38 degrees of freedom were allowed for main sample 1, and 60 degrees of freedom for main sample 2. The hypothesis was considered supported for test I.

The correlations of the variety and novelty measures of test II with the criterion measure of test II, were found to be 0.99 and 0.91 respectively, for main sample 1, and 0.99 and 0.90 respectively for main sample 2. Each of these correlation coefficients was found to be significant at the .05 level, where 38 degrees of freedom were allowed for main sample 1, and 60 for main sample 2. The hypothesis was considered supported for test II.

The correlations of the variety and novelty measures of test 2 with the criterion measure of test 2, were found to be 0.96 and 0.98 respectively, for main sample 1, and 0.98, and 0.96 for main sample 2. Each of these coefficients was found to be significant at the .05

level, where 38 degrees of freedom were allowed for main sample 1, and 60 for main sample 2. The hypothesis was considered supported for test 2.

6.4-11 Discussion -- Hypothesis 9

The results provide some evidence to support the proposition that individual differences in some aspects of inventiveness may be measured by measuring individual differences in facility, variety, and novelty in production. While results from hypotheses 7 and 8 indicate that divergent production leads to inventiveness, the results of the present investigation suggest that measurements of divergent production are significantly related to measurements of inventiveness.

6.4-12 Preliminary Discussion -- Hypothesis 10

Hypothesis 10. That novelty measures for responses, and inventiveness measures for responses correlate significantly.

Preliminary Discussion

The hypothesis was tested by treating standardized responses as "individuals" on which the two observations were made. The novelty score was obtained by statistical and logical methods outlined in chapter 5. The inventiveness measure for each "individual" was the mean of the ratings of the SIGNIFICANT GROUP for that response.

6.4-13 Analysis and Results -- Hypothesis 10

The simple (product-moment) correlations of the two measures for each test administered may be found in table XXXV. The following notation has been used in the table:

Notation	Meaning
r_1	The simple (product-moment) correlation between the novelty and inventiveness measures of the standardized responses of main sample 1
r_2	The simple (product-moment) correlation between the novelty and inventiveness measures of the standardized responses of of main sample 2.
df_1	The degrees of freedom associated with the test of the significance of r_1.
df_2	The degrees of freedom associated with the test of significance of r_2.
NS	An indication that the observed correlation was not significantly different from zero at the .05 level.
S	An indication that the observed correlation coefficient was significantly different from zero at the .05 level.

The hypothesis was not rejected for six of the ten DP tests of main sample 1, and was rejected for the remaining four. The hypothesis was not rejected for seven of the ten DP tests of main sample 2, and was rejected for the remaining three. The particular tests for which these decisions were made are indicated in table XXXV.

6.4-14 Discussion -- Hypothesis 10

The results did not indicate a uniform association between the novelty and inventiveness measures of standardized responses. An attempt was made to interpret the results in terms of the productivity levels of the tests and the product categories in which the tests were classified. Table XXXVI gives the product category of each test, the divergent productivity ratios for each test in terms of main sample 1 (P_1) and main sample 2 (P_2), and the decisions on the significance of correlation coefficients of the novelty and

TABLE XXXV

SUMMARY OF ANALYSIS OF SIGNIFICANCE OF REALTIONSHPS BETWEEN NOVELTY AND INVENTIVENESS MEASURES OF RESPONSE[1]

Test No.	r_1	df_1	D_1	r_2	df_2	D_2
1	0.71	66	S	0.50	50	S
I	0.44	60	S	0.43	33	S
IB	--	--	--	0.65	31	S
1B	--	--	--	0.70	11	S
2	0.45	2	S	0.95	2	S
II	0.96	2	S	0.87	2	NS
IIB	--	--	--	0.64	3	NS
IV	0.59	3	NS	0.58	3	NS
4	0.03	16	NS	0.59	12	S
4B	--	--	--	0.37	62	S
5	-0.15	9	NS	--	--	--
5B	0.61	92	S	--	--	--
VI	0.56	23	S	--	--	--
6B	0.28	42	NS	--	--	--

[1]The notation used is explained in section 6.4-13

inventiveness measures. D_1 and D_2 are the same as in table XXXV. The productivity ratio for a test was the ratio of the number of subjects who produced at least two appropriate responses to the total number of subjects in the sample. The productivity ratio was calculated for Classes as the ratio of the number of students who produced more than one class of responses to the total number of students.

A rearrangement of the tests and decisions in main sample 1, according to descending order of productivity ration indicates a pattern:

Test No.	P_1	D
1	0.98	S
I	0.95	S
VI	0.80	S
5B	0.63	S
6B	0.50	NS
4	0.48	NS
2	(0.45)	S
II	(0.43)	S
VI	0.28	NS
IV	0.15	NS

The pattern indicated is that a significant relationship tends to exist between novelty and inventiveness in responses when productivity is high, but a non-significant relationship tends to exist when productivity is very low.

TABLE XXXVI

NOVELTY-INVENTIVENESS RELATIONSHIPS WITH PRODUCTIVITY LEVELS AND PRODUCT CATEGORIES[1]

Test No.	Product	P_1	D_1	P_2	D_2
1	Units	0.98	S	0.92	S
I	Units	0.95	S	0.87	S
IB	Units	--	-	0.81	S
1B	Units	--	-	0.69	S
II	Classes	(0.43)	S	(0.32)	NS
2	Classes	(0.45)	S	(0.53)	S
IIB	Classes	--	-	(0.10)	NS
IV	Systems	0.15	NS	0.05	NS
4	Systems	0.48	NS	0.29	S
4B	Systems	--	-	0.95	S
5	Trans.	0.28	NS	--	-
5B	Trans.	0.63	S	--	-
VI	Imp.	0.80	S	--	-
6B	Imp.	0.50	NS	--	-

[1]The notation used in this table is explained in section 6.4-14.

A similar pattern is observed for main sample 2 as set out below:

Test No.	P_2	D_2
4B	0.95	S
1	0.92	S
I	0.87	S
IB	0.81	S
1B	0.69	S
2	(0.53)	S
II	(0.32)	NS
4	0.29	S
IIB	(0.10)	NS
IV	0.05	NS

The results may be interpreted as indicating that when productivity was high, the novelty of a response was a significant predictor of its inventiveness, but when productivity was very low, novelty was insufficient to account for inventiveness. The interpretation is similar to that given for the result of the hierarchical analysis of the DP measures within product categories of hypothesis 5. It was suggested that where productivity was high, novelty was the most complex quality of the three DP measures, but when productivity was very low, novelty seemed the most basic quality. Although in this case reference is made to novelty in responses, the results appear comparable.

A similar point of view depends on the observation that a uniform significant relationship was found among the measures of responses in the Units category, but no uniform result was found in the other categories. It may be considered that where the task lends

itself easily to the production of elementary ideas, the most inventive ideas tend to be the most rare ones in the population, but where few ideas seem possible, rarity is insufficient to account for the most inventive responses.

6.5 SUMMARY AND CONCLUSION

Analytical investigations have been conducted on content, construct, and criterion validation of tests designed to be tests of inventiveness in school mathematics. The tests were divergent production (DP) tests in mathematics, and they were primarily designed to measure abilities defined in terms of some of the psychological products which formed the components of one of parameters in Guilford's structure-of-intellect model.

The propositions which formed the foundations of the study were:

(i) that it was possible to evoke and measure inventiveness in mathematics,

(ii) that divergent production in mathematics would lead to inventiveness in mathematics, and

(iii) that individual differences of students in some aspects of inventiveness in mathematics may be ascertained from their individual differences in the components of divergent production in mathematics.

The investigation on the content validity of the tests was conducted by obtaining the assessments of experts in mathematics, mathematics education, and measurement, on the face and content validities of the tests. The investigations on construct validity took the form of ascertaining to what extent certain vital expectations arising from the theoretical bases of the tests were experimentally met. These expectations concerned the extent to which the tests were suitably classified in product categories according to the

abilities defined for products in that category, the extent to which the tests related to subject mastery and problem solving, and the extent to which certain logical expectations of structural hierarchy among the products were met. The investigations on criterion validity related to the extent to which the tests evoked and measured inventiveness.

It was found that the experimental evidence did not contradict the propositions which formed the foundations of the study. Fourteen DP tests were tried out in two main samples, and six of the tests were common to both samples. It was found that all fourteen tests evoked inventiveness in students. All ten of the DP tests of main sample 1, and nine out of ten of the tests of main sample 2 were found to have been suitable for some product-measure ability, where the "products" were Units, Classes, Systems, Transformations, and Implications, and the "measures" were Facility, Variety, and Novelty in production. It was found that nine of the ten tests of main sample 1, and four of the ten tests of main sample 2, were "stable" over the DP measures, in the sense that they were suitable tests for all the product-measure categories in which they were used. It was also found that in the case of at least four of the tests the DP measures predicted a criterion measure of inventiveness significantly.

It is concluded that the propositions which formed the foundations of the study have been supported by empirical evidence, and that all fourteen DP tests have validity as tests of inventiveness, but that the "stable" tests have met the most severe expectations for validity.

CHAPTER VII

DISCUSSION, CONCLUSIONS, AND IMPLICATIONS

7.1 INTRODUCTION AND OVERVIEW

The problem of the study was to establish principles for constructing and scoring tests of inventiveness in mathematics at the senior high school level. Theoretical and experimental methods were used to establish these principles. A review of literature in the second chapter of this dissertation led to the statement of Guilford (1965, p. 15), that "most of the more obvious contributions to creative thinking are in the divergent production category." This statement was interpreted as a proposition that divergent production leads to inventiveness. The principles for constructing and scoring tests of inventiveness adopted in this study were based on three propositions which were formulated for the identification and measurement of inventiveness in school mathematics. These three fundamental propositions were stated in the second chapter, and were based on the proposition that divergent production leads to inventiveness. Divergent production and its relation to inventiveness was discussed in the third chapter, and principles for constructing divergent production tests to evoke inventiveness in mathematics were outlined in the fourth chapter. The principles for scoring the tests were set out in the fourth chapter. Theoretical developments on the basic scoring principles were made in the fifth chapter.

The experimental establishment of the principles consisted of validating tests of inventiveness in mathematics constructed and

scored using the formulated principles of divergent production in mathematics. Analytical investigations on the validities of the tests were reported in the sixth chapter. It was established that all the fourteen tests of the main studies were valid as tests of inventiveness in high school mathematics, although it was found that some had satisfied more critical expectations for validity than others. It was concluded in the sixth chapter that the three propositions which formed the foundations of the study had been supported by empirical evidence. The establishment of the validity of all the tests of the main studies as tests of inventiveness in high school mathematics, constituted the establishment of the principles of the study.

Some important aspects of the study are discussed in this closing chapter. Further comment is provided on the fundamental propositions of the study, the principles of the study, and on Guilford's structure-of-intellect theory as applied to divergent production in school mathematics. Observed and inferred characteristics of the tests are argued and highlighted. Uses of divergent production approaches to achieve the objective of inventiveness in classroom mathematics are discussed, and a particular discovery-teaching method is modified to include techniques of divergent production in mathematics, as a way of embedding techniques of divergent production into a suitable teaching method. The chapter closes with some problems for further research, and a brief epilogue.

7.2 FUNDAMENTAL PROPOSITIONS

The first of three fundamental propositions states that "it is possible to evoke and measure inventiveness in mathematics."

This is a very important proposition. Methods of measuring inventiveness in mathematics deserve consideration only if this proposition is valid. The validity of this proposition is inferred from evidence that the tests investigated in this study measure inventiveness.

It is important to stress that while the method of measuring inventiveness in this study has been through divergent production tests, the validity of the first proposition could have been established through other possible methods of measuring inventiveness in mathematics, as for example, through convergent problem solving mathematics tests, or through teacher and peer ratings of inventiveness in mathematics.

The second proposition states "that divergent production in mathematics leads to inventive production in mathematics." The evidence shows that all fourteen divergent production tests administered in the main studies evoked inventive productions, and that each subject in each of the main samples produced at least one inventive response. The evidence indicates that the process of divergent production results in inventive productions

The third proposition states "that individual differences of students in some aspects of inventiveness in mathematics may be ascertained from their individual differences in the components of divergent production in mathematics." The components of divergent production are facility, variety, and novelty in production. Evidence has been presented to show that in the case of at least four tests, the measures of each of the components correlate significantly with the criterion measure of inventiveness. This establishes the validity of the third proposition.

The establishment of the second and third propositions imply the establishment of the first proposition. It is concluded that methods of divergent production in mathematics may be used to provide valid principles for constructing and scoring tests of inventiveness in mathematics.

7.3 THE PRINCIPLES OF THE STUDY

The principles for constructing and scoring tests of inventiveness in high school mathematics are based on the three fundamental propositions of the study. They are rules for guidance in constructing and scoring tests of inventiveness, and represent an elaboration of the three fundamental propositions of the study, and an application of these propositions to the classroom situation. The principles established in this study for constructing tests of inventiveness in high school mathematics are given below (with some generalizations).

1. Each test should present a mathematical situation, and request more than one response production to the situation.

2. Several appropriate responses should be conceivably possible.

3. Response productions of high quality should be conceivably possible.

4. Each test should be subject-related. This means that the basic subject matter of each test should be based on the mathematics taught in class.

5. Each test should be such that a student who knows the subject matter should be capable of making at least one appropriate response production.

6. Each test should be designed to measure a particular divergent production (DP) ability of production in mathematics.

The first principle is designed to ensure that students are requested to engage in divergent production.

The second principle is a guiding rule for the test constructor. It should in his judgement be possible for several appropriate response productions to be made to the mathematical problem situation. The test constructor could effectively ensure this by producing several appropriate responses himself.

The third principle is designed to ensure that in the judgement of the test constructor, each test is capable of evoking inventive responses.

The fourth principle requires the content of the tests to be based on the classroom mathematics. Prima facie non-classroom mathematics which could be conceivably tackled using classroom mathematics may also be included. This principle is in accordance with the conception of subject-related inventiveness in mathematics as part of achievement in mathematics.

The fifth principle requires the test constructor to ensure that in his judgement every student who knows the subject matter of the test could make an appropriate response production. It is recommended that students be allowed to consult their text books during the tests.

The sixth principle requires the test constructor to aim at testing a particular DP ability. It has been shown in this study that this can be done. Constructing tests to measure a particular ability is an effective procedure in constructing valid tests.

The DP abilities of production in mathematics established in the main studies are

(i) The ability to produce various elementary mathematical ideas related to a mathematical situation.

(ii) The ability to resist fixedness in mathematical thinking, and to produce mathematical ideas that are different in relation to a mathematical situation.

(iii) The ability to organize elementary mathematical ideas into complex ones.

(iv) The ability to produce original mathematical ideas involving re-interpretations and redefinitions.

(v) The ability to produce mathematical implications from a given set of conditions.

The definitions of the above DP abilities were adapted from definitions of DP abilities given by Guilford and Hoepner (1966).

The DP mathematical abilities were conceived as abilities of production resulting in mathematical products. These products correspond to the products of Guilford's structure-of-intellect model (Guilford, 1967). Guilford's six products are Units, Classes, Relations, Systems, Transformations, and Implications. The five DP mathematical abilities investigated in the main studies are abilities defined in terms of five of Guilford's six products. They may be referred to as abilities of

(i) Divergent production of mathematical units (DMaU),

(ii) Divergent production of mathematical classes (DMaC),

(iii) Divergent production of mathematical systems (DMaS),

(iv) Divergent production of mathematical transformations (DMaT), and

(v) Divergent production of mathematical implications (DMaI).

The principles for scoring DP mathematics tests established in this study are

1. DMaC tests should be marked for variety and novelty in production.

2. DMaU, DMaS, DMaT, and DMaI tests should be marked for facility, variety, and novelty in production.

The facility score of a subject in a DP test is the number of non-identical response productions he makes, which satisfy the requirements of the test.

The variety score of a subject in a DP test is the number of "different" response productions he makes, where the determination of "different" depends on the procedure adopted by the examiner.

The novelty score for a subject in a DP test is the average of the two highest novelty scores assigned to his response productions in the test. The novelty score assigned to a response production represents the degree of rarity of the response production in relation to all the response productions of all the subjects in the test.

Certain developments were made in this study to reduce the subjective aspects of the determination of the variety score, and to set up some recommended procedural rules for determining the variety score. Two variety scores were developed, a <u>representative concepts variety score</u>, and a <u>specific concepts variety score</u>.

In developing the variety scores, certain notions were developed leading to the notion of a <u>representative</u> of the <u>concept set</u> of a response production. Such a representative may be loosely described in terms of the response production. The representative of a response production would correspond to a response production which totally embodies the important mathematical qualities of the original response production, but is itself not totally embodied by any other response production in the collection of responses.

More formally, the notion of the <u>criterion mathematical concept</u> of a response production was developed. The criterion was a quality in response productions which was chosen as important in making

distinctions among them. The subject matter domain of responses was chosen as such a quality, and "the concept of the subject matter domain" of response productions was chosen as the criterion mathematical concept of the response productions of this study.

The notion of the concept set of a response production was developed as the set of all mathematical properties necessary to establish the criterion mathematical concept of the response production. The representative of a concept set was that concept set which contained it, but was contained in no other concept set in the universe of all concept sets being considered.

The idea of a maximally reduced subuniverse of a universe of concept sets was introduced, and it was found that it was the same as the set of all representatives of a universe containing a finite number of distinct concept sets. It was established that there was one and only one maximally reduced subuniverse for any universe containing a finite number of distinct concept sets. This established the existence and uniqueness of the number of representatives of the concept sets of a universe containing a finite number of distinct concept sets.

The representative concepts variety score was the number of representatives in a universe of concept sets, and the specific concepts variety score was the number of representatives which were not contained in the totality (union) of the other representatives.

The investigator recommends that where the relative individual differences of subjects are required as was the case in this study, representatives should be determined from the totality of response productions made by all the students of the sample. These representatives become the sample representatives. The variety score for each subject

should be the number of sample representatives of the concept sets of his response productions. The following procedure is recommended for obtaining the maximally reduced subuniverse (set of all representatives) of the universe of concept sets of response productions.

Stage	Procedure
1	Obtain the concept set for each response production.
2.	Ensure that only distinct concept sets are retained.
3.	Successively discard any concept set that is contained in another, doing this by logical analysis of mathematical relationships. Obtain a <u>possible maximally reduced subuniverse</u>.
4.	Verify whether the obtained possible maximally reduced subuniverse is the maximally reduced subuniverse using a two dimensional matrix table, in which the ith row and jth column position in the matrix of results contains the answer to the question "Is the ith concept set a subset of the jth concept set?" An affirmative result is indicated by a 1, and a negative result by a 0. The maximally reduced subuniverse is obtained if, and only if, the resulting square matrix contains ones in the diagonal, and zeros everywhere else.

The Guilford notion of using rarity of responses to determine the novelty score for a response has been adapted to take consideration of situations in which trivial responses are as rare as responses of quality. The procedure recommended for obtaining the novelty score of response productions is to assign upper bound scores to the response productions according to the proportion of subjects making the response

production. A response production Q is awarded its upper bound score if, and only if, every response production in the sample of all response productions, in whose concept set Q is contained, has an upper bound score greater than or equal to the upper bound score of Q. Otherwise, its upper bound score is reduced to the upper bound score of a response production whose concept set contains the concept set of Q. The procedure is continued until each response can be awarded its upper bound score.

7.4 GUILFORD's STRUCTURE-OF-INTELLECT THEORY AS APPLIED TO DIVERGENT PRODUCTION IN SCHOOL MATHEMATICS

The study has attempted to apply the divergent production aspects of Guilford's structure-of-intellect model to achievement in school mathematics. "Divergent Production" is one of the "operations" in Guilford's model, and associated with each operation are four kinds of content (figural, symbolic, semantic, and behavioral), and six kinds of products (units, classes, relations, systems, transformations, and implications).

Guilford's theory is basically concerned with "The nature of human intelligence," and creativity is considered by Guilford as part of intelligence. The present study is concerned with inventiveness as part of school achievement. Certain adaptations were considered essential in the application of Guilford's theory to this study, to accord with the conception of subject-related inventiveness. A student's response production to a divergent production test was expected to involve some or all of the components of content in Guilford's model, and no distinction was made in this study among the kinds of content. However, distinctions were made in terms of the six kinds

of products.

The mathematical abilities investigated were defined in terms of production in Guilford's product categories. Tests measuring abilities in each of the six product categories were investigated in a preliminary study. The tests constructed by the investigator for the "relations" category did not meet the criteria for final selection (Appendix B).

The results of the investigations concerned with the construct validation of the tests, may be used to discuss Guilford's theory as it applies to mathematics. The factorial investigations indicated that abilities defined in terms of units, classes, systems, transformations, and implications, had some valid claims to existence, although not necessarily to independent existence.

The hierarchical investigations provided an important verification of the "logical order" suggested by Guilford for each of the dimensions of his model. When investigations were carried out on the product categories for each of the three components of divergent production (facility, variety, novelty), no general hierarchy was found connecting units, classes, systems, transformations, and implications. However, hierarchical linkages were found among several groups of three of these products, and they accorded with Guilford's "logical order," except in the case involving facility in implications. It was found that facility in implications was less hierarchically complex than facility in units, but variety in implications was more hierarchically complex than variety in classes, as well as variety in systems. This suggests that the hierarchical complexity of implications depends on the measure being considered. Guilford also had

some reservations on the position of implications (1967, p. 63).

The findings that the hierarchical analyses of this study tended to support Guilford's "logical order" is of some importance in discussing the validation of Guilford's model as applied to mathematics. The investigator has not been able to find any study in which the hierarchical nature of the dimensions of Guilford's model has been investigated. The statistical procedures developed by the investigator for this purpose were based on developments by Guttman (1954), Burt (1954), and Moise (1963), and may be used for further similar investigations.

7.5 OBSERVED AND INFERRED CHARACTERISTICS OF TESTS OF DP MATHEMATICAL ABILITIES

7.5-1 Divergent Production of Mathematical Units (DMaU)

The four DMaU tests administered in the main investigations were tests I, 1, IB, and 1B of section 4.2-2. Tests I and 1 were administered to both main samples, and tests IB and 1B were administered to main sample 2 only. Each test was found to evoke inventive response productions (Table XXXIV).

A breakdown is given below of the DMaU tests in terms of their being found "suitable" or "stable" on the basis of the factor analytic studies (section 6.3). "Suitable" tests were tests which had been hypothetically classified in the same product category, and were experimentally found to determine the same construct for a DP measure. "Stable" tests were tests which were "suitable" for all measures for which they were investigated. Productivity ratios (PR) are included below for easy reference. The situation for main sample 1 was as follows:

Test No.	Measure Suitability Domains	Status	PR
I	Facility, Variety, Novelty	Stable	0.98
1	Facility, Variety, Novelty	Stable	0.95

and the situation for main sample 2 was as follows:

Test No.	Measure Suitability Domains	Status	PR
I	Facility, Variety, Novelty	Stable	0.92
1	Facility, Variety, Novelty	Stable	0.87
IB	Facility, Novelty with tests I and 1, Variety with test 1B.	Unstable	0.81
1B	Variety, with test IB.	Unstable	0.69

The DMaU tests were designed to measure "the ability to produce various elementary mathematical ideas related to a mathematical situation." A possible reason why test 1B failed to be a good DMaU test is that the mathematical function given to the students in this test was too unfamiliar to lead to "elementary" productions. This test was found to load significantly on the same facility factor as the two systems tests IV and 4.

A striking characteristic of the DMaU tests is the high productivity of the "suitable" tests, and the particularly high productivity of the "stable" tests.

The results of the analytical investigations in connection with Hypothesis 10 revealed that the novelty-inventiveness relationships of response productions for each DMaU test was significant. The novelty measure of response productions was obtained from statistical and logical considerations, and was an indication of the rarity of the response productions. The inventiveness measures of response productions was their average assessments by the SIGNIFICANT GROUP.

Although novelty is generally considered to be closely related to inventiveness or creativity, the relationship is not generally considered to be perfect. Stein's definition of a creative work includes not only the requirement that the work should be novel, but that it should be "accepted as tenable or useful or satisfying to a significant group of others at some point in time." Whitting (1956, p. 3) considers that "an original idea that is also useful in terms of meeting some of man's needs, is also a creative idea." Bruner (1967) who defines creativity as "effective surprise -- the production of novelty" (p. 28), expects more than rarity to result in effective surprise. He distinguishes between "trivial improbabilities" and "effective surprise," and asserts that it takes preparation to discern the difference (1967, p. 4).

The significant novelty-inventiveness relationship of DMaU response productions may be interpreted as indicating that the quality of rarity in DMaU response productions is significantly associated with those qualities which make the response productions identifiable as inventive.

The basic principle used to construct the DMaU tests, and the observed and inferred characteristics of the tests are as follows:

<u>Basic Principle</u>

Each test should be designed to measure the ability to produce various elementary mathematical ideas related to a mathematical situation.

<u>Characteristics</u>

1. Each test evokes inventive responses.

2. Productivity ratios are generally high ($\geq .61$), and

particularly high ($\geqslant .81$) for the stable tests.

3. The novelty-inventiveness relationship of response productions is significant for each test.

7.5-2 Divergent Production of Mathematical Classes (DMaC)

The three DMaC tests administered in the main investigations were tests II, 2, and IIB of section 4.2-2. Tests II and 2 were administered to both main samples, and test IIB was administered to main sample 2 only. Each test was found to evoke inventive response productions (Table XXXIV).

A breakdown is given below of the DMaC tests in terms of suitability and stability. Productivity ratios (PR) are included for easy reference. The situation for main sample 1 was as follows:

Test No.	Measure Suitability Domains	Status	PR
II	Variety, Novelty	Stable	0.43
2	Variety, Novelty	Stable	0.45

and the situation for main sample 2 was as follows:

Test No.	Measure Suitability Domains	Status	PR
II	Variety, Novelty	Stable	0.32
2	Variety, Novelty	Stable	0.53
IIB	Variety	Unstable	0.10

The DMaC tests were designed to measure "the ability to resist fixedness in mathematical thinking and to produce mathematical ideas that are different in relation to a mathematical situation." The instructions to these tests requested the students to produce ideas that were similar, group them together, strive for a different pattern of ideas, and group those together. The students were to strive for a large number of different patterns of ideas. The

variety score in this case was a measure of a flexibility which was deliberate and not spontaneous.

The productivity ratio for the unstable test was low (.10), and the productivity ratios for the stable tests were moderate.

The novelty-inventiveness relationship of response productions was found to be significant for the two stable tests of main sample 1, but it was only for test 2 of main sample 2 that the relationship was found to be significant. The observation that the significant novelty-inventiveness relationship occurred with the test with the highest DMaC productivity ratio in main sample 2, is in accordance with the following hypothesis formulated a posteriori.

<u>A Posteriori Novelty-Inventiveness Hypothesis</u>

> In a given sample, the productivity ratio of a test whose response productions have a significant novelty-inventiveness relationship, is always greater than the productivity ratio of any test in the same product category whose novelty-inventiveness relationship is non-significant.

The novelty-inventiveness relationship of response productions of tests which are in accordance with the above hypothesis, may be interpreted as indicating that for a test with relatively high productivity ratio within a product category in a particular sample, the quality of rarity in the response productions to that test is significantly associated with those qualities which make the response productions identifiable as inventive. However, for a test with relatively low productivity ratio, the quality of rarity in the response productions to that test is not significantly associated with those qualities which enable the response productions to be identifiable as inventive.

The basic principle used to construct the DMaC tests, and the observed and inferred characteristics of the tests are as follows:

Basic Principle

Each test should be designed to measure the ability to resist fixedness in mathematical thinking, and to produce mathematical ideas that are different in relation to a mathematical situation.

Characteristics

1. Each test evokes inventive response productions.

2. Productivity ratios are low (.10) and moderate (0.32-0.53). Productivity ratios for stable tests are moderate.

3. The novelty-inventiveness relationships of DMaC tests are in accordance with the "A posteriori novelty-inventiveness hypothesis" of this section (section 7.5-2).

7.5-3 Divergent Production of Mathematical Systems (DMaS)

The three DMaS tests administered in the main investigations were tests IV, 4, and 4B of section 4.2-2. Tests IV and 4 were administered to both samples, and test 4B was administered to sample 2 only. Each test was found to evoke inventive responses (Table XXXIV).

A breakdown is given below of the DMaS tests in terms of suitability and stability. Productivity ratios (PR) are included for easy reference. The situation for main sample 1 was as follows:

Test No.	Measure Suitability Domains	Status	PR
IV	Facility, Variety, Novelty	Stable	0.15
4	Facility, Variety, Novelty	Stable	0.48

and the situation for main sample 2 was as follows:

Test No.	Measure Suitability Domains	Status	PR
IV	Facility	Unstable	0.05
4	Facility	Unstable	0.29
4B	(None)	Unstable	0.95

The DMaS tests were designed to measure "the ability to organize elementary mathematical ideas into complex ones." The tests were expected to lead to the production of complex ideas. A possible reason why test 4B failed to be a good DMaS test is that the task of inventing complex and unusual sequences as requested in test 4B, did not prove to be a complex task. The high productivity ratio of this test (.95) suggests that production was easy as might be expected of DMaU tests, rather than complex and rather difficult as would be expected from DMaS tests. Further evidence of close association of test 4B with DMaU tests is provided by the results of the factorial investigations. Test 4B loaded significantly on the same facility factor as did DMaU tests I, 1, and IB, on the same variety factor as DMaU tests IB and 1B, and on the same novelty factor as did DMaU test IB.

Test IV and 4 were found to be stable in main sample 1, but unstable in main sample 2. Every subject in main sample 2 scored zero for variety in test IV, and this test was not included in the factorial analysis of the variety variables of main sample 2. This was a contributory cause to the instability of this test for main sample 2. The very low productivity ratio of this test (.05) may be too low for a DP test to show stability in a sample.

It was observed that the suitable DMaS tests had low (.05 and .15) and low-moderate (.29 and .48) productivity ratios.

The novelty-inventiveness relationship of response productions of DMaS tests (see table XXXVI) of sample 1 were both non-significant. This observation did not contradict the "A posteriori novelty-inventiveness hypothesis" of section 7.5-2.

The novelty-inventiveness relationship of tests 4 and 4B were found to be significant in main sample 2, but the relationship was found to be non-significant for test IV. This finding was in accordance with the "A posteriori novelty-inventiveness hypothesis" of section 7.5-2.

The basic principle used to construct the DMaS tests and the observed and inferred characteristics of suitable DMaS tests are as follows:

Basic Principle

Each test should be designed to measure the ability to produce complex mathematical ideas related to a mathematical situation.

Characteristics

1. Each test evokes inventive responses.
2. Productivity ratios are low and low-moderate.
3. The novelty-inventiveness relationships of DMaS tests are in accordance with the "A posteriori novelty-inventiveness hypothesis" of section 7.5-2.

7.5-4 Divergent Production of Mathematical Transformations (DMaT)

Two DMaT tests were administered in the main investigations to main sample 1 only. These were tests 5 and 5B. Each test was found to evoke inventive responses (Table XXXIV.)

A breakdown of the DMaT tests in terms of suitability and stability is given below. Productivity ratios (PR) are included for easy reference.

Test No.	Measure Suitability Domains	Status	PR
5	Facility, Variety, Novelty	Stable	0.28
5B	Variety, Novelty	Unstable	0.63

The stable DMaT test shows a low-moderate productivity ratio, and the unstable DMaT test shows a high-moderate productivity ratio. It is perhaps due to the rather high productivity ratio of test 5B that it proved unstable. This test loaded significantly on the same facility factor as did the stable units tests I and 1.

The DMaT tests were designed to measure "the ability to produce original mathematical responses involving re-interpretations and redefinitions." A deliberate effort was made to ensure that the mathematical tasks necessitated the shifting of ideas from one frame of reference to another frame of reference. It would appear that when this can be easily done, the "facility" results have much in common with production of units, as was the case with test 5B.

The novelty-inventiveness relationship of DMaT response productions investigated in Hypothesis 10 revealed a non-significant relationship in test 5 where productivity was low-moderate (.28) and a significant relationship in test 5B where productivity was high-moderate (.63). The trend of the relationships was in accordance with the "A posteriori hypothesis" of section 7.5-2.

The basic principle used to construct the DMaT tests, and the observed and inferred characteristics of suitable DMaT tests are as follows:

Basic Principle

Each test should be designed to measure the ability to produce original mathematical responses involving re-interpretations and redefinitions.

Characteristics

1. Each test evokes inventive responses.

2. Productivity ratios are moderate.

3. The novelty-inventiveness relationships of response productions are in accordance with the "A posteriori novelty-inventiveness hypothesis" of section 7.5-2.

7.5-5 Divergent Production of Mathematical Implications (DMaI)

Two DMaI tests were administered in the main studies to main sample 1 only. These were tests VI and 6B. Each test was found to evoke inventive responses (Table XXXIV).

A breakdown of the DMaI tests in terms of suitability and stability is given below. Productivity ratios (PR) are included for easy reference.

Test No.	Measure Suitability Domains	Status	PR
VI	Facility, Variety, Novelty	Stable	0.80
6B	Facility, Variety, Novelty	Stable	0.50

The DMaI tests were designed to measure "the ability to produce mathematical implications from a given set of conditions." An attempt was made to set up conditions which were prima facie unfamiliar to the students, but such that the mathematical background of the students should enable them to cope with the task of producing implications to these conditions. The high productivity ratio of test VI is suggestive of units tests. However, the DMaI tests were found to be factorially associated with only the facility measure for the DMaU test 1, in terms of determining the same factor.

The novelty-inventiveness relationship of DMaI response productions investigated in Hypothesis 10 revealed a non-significant relationship for test 6B which had the lower productivity ratio (.50), and a significant relationship for test VI which had the higher productivity ratio. The trend of the relationship was in accordance

with the "A posteriori novelty-inventiveness hypothesis" of section 7.5-2.

The basic principle used to construct the DMaI tests and the observed and inferred characteristics of the tests are as follows:

Basic Principle

Each test should be designed to measure the ability to produce mathematical implications from a given set of conditions.

Characteristics

1. Each test evokes inventive responses.
2. Productivity ratios are high-moderate and high.
3. The novelty-inventiveness relationships are in accordance with the "A posteriori novelty-inventiveness hypothesis" of section 7.5-2.

7.5-3 General Observations

The tests for each DP mathematical ability evoked inventive response productions. Where interest lies in evoking inventiveness, tests of any of the abilities are suitable.

The productivity ratios for suitable DMaU and DMaI tests tended to be high-moderate to high. The productivity ratios for the suitable DMaC, DMaS, and DMaT tests tended to be low to high-moderate. Where interest lies in evoking inventive responses from as many students as possible, DMaU and DMaI tests are the most suitable. Tests of the other abilities may be suited to fine discrimination among highly inventive subjects.

There is evidence to suggest that where productivity is very

high in a sample, the quality of rarity in response productions is significantly associated with those qualities which enable the response productions to be identifiable as inventive, but that where productivity is very low, the quality of rarity in response productions is not significantly associated with those qualities which enable them to be identifiable as inventive.

7.6 CLASSROOM USES OF DIVERGENT PRODUCTION

7.6-1 General Classroom Uses

Since DP mathematics tests evoke inventive responses, they may be used to advantage for this purpose in the mathematics classroom. One use would be for a teacher to use DMaI tests to lead students to consider mathematical situations which had not yet been presented in class. This would give the students the opportunity of formulating mathematical ideas for themselves. Students may be able to perceive and invent relationships under such conditions.

Another use would be to use DMaU tests on a mathematical topic which had already been presented in class. This would provide the teacher with some information on what the students mastered from his instruction, and what new insights the students had on what he had taught.

DP mathematics tests may be used with convergent mathematics problems. They may be used to lead students to produce a variety of ideas on aspects of the problem, to formulate similar problems, to strive for different methods of solving the problem, and to restate the problem so that it could be understood by groups of other students with different but specified mathematical capabilities, as for example mathematics students who had mastered grade nine mathematics.

DP measuring procedures may also be used to evaluate achievement in school mathematics, where achievement is concerned not only with subject mastery and the reproduction of mathematics, but also with inventiveness, and the production of mathematics. The procedures would provide an added dimension to the usual evaluative procedures of school mathematics.

DP testing procedures used in the process of teaching, may inspire confidence in mathematics. Students may find themselves making response productions which may surprise and delight them. Both student and teacher may come to appreciate additional potentialities in the student.

7.6-2 The Mathematizing Mode

An example of a teaching method and its modified form using divergent production approaches is given below. This is presented as an example of embedding divergent production approaches into an appropriate teaching method.

The teaching method that is modified is a form of discovery method developed by Sigurdson (1968) and Johnston (1968), and is called the Mathematizing Mode.

The mathematizing mode has four main stages:

Stage 1. A period of uninhibited exploration of a problem situation on the part of the pupils.

Stage 2. A period of "brainstorming" in which the teacher acts as a moderator and a scribe. Each suggestion and attempt at a hypothesis is accepted without evaluation. Every attempt at a hypothesis is accepted without evaluation. Every statement by the pupils is to be rewarded by the teacher as contributing to the learning process.

Stage 3. Hypothesis Testing. The teacher initiates questions and problems which will enable students to test out their hypotheses.

Stage 4. Summing Up. This is a period of summing up of the precise mathematical principles involved in the preceding stages. Here the student is made fully aware of the present day conventions and language.

It is suggested that an adapted form could be as follows:

Stage 1. Presentation. The teacher presents a mathematical problem situation to the students and requests them to make written divergent response productions to the situation.

Stage 2. Student marking of the responses of each other. Each student marks the papers of each other in terms of facility, variety and novelty, the teacher providing guidance. All productions which could not be assessed immediately as right or wrong should be considered as "possibilities" and reserved for further investigation.

Stage 3. The investigation of possibilities and the testing of hypotheses. The possibilities of stage 2 should now be further investigated. Intuitive responses may now be rigorously investigated. Partially true statements may be given appropriate boundary conditions. Some of the possibilities would be hypotheses. Other hypotheses suggested by the students and the teacher should now be investigated. The teacher should introduce ideas which he feels should be included.

Stage 4. Same as stage 4 of the mathematizing mode.

It is perhaps, of importance here to give examples of possible responses that may be considered as "possibilities" of stage 2. The "possibilities" given below are actual or adapted statements made by students in the study.

Test 1. Write down as many mathematically true statements as you can about an Epudom in the sense defined below:

An Epudom is an integer divisible by 35.

Possibility: (Actual Response) The sum of the first and last digits in an Epudom never exceeds 14.

Test 1B. Write down as many mathematically true statements as you can about the following function. Each statement should be true or

would be true under certain conditions. Try to make each statement represent one idea only. Use your imagination.

$$y = 2^x, \quad x \in R.$$

Possibility: (Actual Response) x and y can never be equal.

Test IV. Show that if b and c are real, and x_1 and x_2 are the roots of the quadratic equation $x^2 + bx + c = 0$, then $x_1 + x_2 = -b$, using as many distinct methods as you can. When you have thought out a method, write down an outline of the method, such that it would be possible to see how you would proceed if you had time. Then, look for another method. Try to think out and write down outlines of as many methods as you can.

Possibility: (Adapted)

Substituting the roots in the equation and subtracting we obtain $x_1^2 + bx_1 + c = x_2^2 + bx_2 + c$. Hence $x_1^2 - x_2^2 = -b(x_1 - x_2)$. The result follows after dividing by $x_1 - x_2$, when x_1 is not the same as x_2.

Test 5. Imagine that you wish to explain to Grade Six students why we cannot divide by zero. Think out some unusual mathematical approaches that you can use. List as many of these approaches as you can.

Possibility: (Actual Response) Take a piece of pie. Cutting once means dividing by 2 (2 resulting pieces). Cutting 3 times means dividing by 3 (3 pieces). Dividing by one means no cutting. There we are at the end of the line. Continuing the analogy dividing by zero means taking less than no cutting. This is ridiculous.

Test 6B. T(n) is defined as the set of all positive integers less than n which divide n evenly. Think out and guess as many properties of T(n) as you can.

Possibility: (Actual Response). The sum of all members of T(n) is less than 2n.

It is a problem for further research to investigate whether the above would be an effective teaching method.

7.7 PROBLEMS FOR FURTHER RESEARCH

A problem of critical importance in the measurement of inventiveness is the determination of a measurable criterion for inventiveness. The position taken in this dissertation is that inventiveness in productions is effectively a quality attributed to them by informed society. General society accepts a production as inventive if informed society accepts it as inventive. The SIGNIFICANT GROUP used in this study is thus expected to be a sample of informed society. There is need to conduct a similar study with the purpose of investigating to what extent different samples of informed society give similar assessments. The high agreement among the SIGNIFICANT GROUP is striking and suggestive of generalizability. It may be that mathematics productions lend themselves to considerable agreement among informed judges of their quality.

An attempt has been made to reduce the subjective aspects of the score for flexibility. However it is still necessary to develop some aspects subjectively. It is a problem for further research to determine whether the subjective aspects of the flexibility score could be further reduced, and whether the present methods significantly reduce this subjectivity.

An attempt has been made to account for the rare but trivial production in assessing the novelty score. The development used by the investigator needs to be further investigated on its empirical validity.

Problem solving is considered by many as the essence of inventiveness, and methods of measuring convergent problem solving may be another way of measuring aspects of inventiveness. It is a problem for further research to determine to what extent measures of inventiveness in mathematics predict other valid methods of measuring inventiveness in mathematics.

Some of the problems discussed above and others are listed below as problems for further research.

1. How consistent over samples are expert ratings of the inventiveness of response productions?
2. To what extent is the novelty-inventiveness relationship of response productions to a test related to the productivity ratio of the test?
3. How could the subjective aspects of the score for flexibility be further reduced?
4. How empirically valid are the variety and novelty scoring procedures reported in this dissertation?
5. To what extent does divergent production in mathematics correlate with other methods of inventiveness in mathematics?
6. To what extent can convergent problem ability significantly predict creative productivity?
7. How reliable can students be as examiners of DP tests?
8. To what extent is the ability to recognize correct answers

as in multiple choice techniques, a capable predictor of the ability to produce significant mathematics?

9. What is the relative effectiveness of the mathematizing mode with adaptations involving DP methods, in relation to other methods, in regard to criteria of interest as for example:

(i) Attitude to mathematics.

(ii) Non-routine problem solving.

(iii) Achievement emphasizing knowledge and comprehension.

(iv) Achievement emphasizing analysis, synthesis, and evaluation.

(v) Originality.

(vi) Flexibility.

7.8 EPILOGUE

The underlying purpose of this study was to establish that it is possible to measure inventiveness in school mathematics and to adopt methods which would lead to effective measurement of inventiveness in the senior high school.

The methods used in this study to evoke and measure inventiveness are methods of divergent production, and evidence has been collected to show that divergent production in mathematics leads to inventiveness in mathematics.

The use of divergent production has been advocated in the classroom, and suggestions have been made of how divergent production may be used in the classroom.

If it is accepted that the study of mathematics should lead to the ability to invent mathematics, then there is a need to make this aim a practical classroom reality. It is the opinion of the investigator that there is such a need, and that this need is pressing.

It is hoped that this dissertation would go some way to providing foundations for the practical outworking of the translation of this need into the reality and adventure of the mathematics classroom.

BIBLIOGRAPHY

BIBLIOGRAPHY

Allen, Myron S. Morphological creativity: The miracle of your hidden brain power. Englewood Cliffs, N. J.: Prentice-Hall, 1962.

Anastasi, Anne. Psychological testing. New York: Macmillan, 1961.

APA-AERA-NCME (approved). Standards for educational and psychological tests and manuals . In Douglas N. Jackson & Samuel Messick (Eds.), Problems in human assessment. New York: McGraw-Hill, 1967.

Atkinson, Richard C., et al. An introduction to mathematical learning theory. New York: John Wiley & Sons, Inc., 1965.

Ausubel, David P. Educational psychology: A cognitive view. New York: Holt, Rinehart & Winston, Inc., 1968.

Barron, Frank. Participant in Process versus product: A spontaneous discussion by conference participants . In Calvin W. Taylor (Ed.), Widening horizons in creativity. New York: John Wiley & Sons.

Bechtoldt, Harold P. Construct validity: A critique. The American Psychologist, 1959, 619-629.

Beesack, P. R., et al. Secondary school mathematics Grade Eleven. Vancouver: The Copp Clark Publishing Company, 1962.

Biggs, E. E. The teaching of mathematics at primary level, Mathematics in commonwealth schools. Report of a Specialist Conference held at the University of the West Indies, St. Augustine, Trinidad, September, 1968. London: Commonwealth Secretariat, 1969.

Bruner, Jerome S. The conditions of creativity. In Glen Terrell & Michael Wertheimer (Eds.), Contemporary approaches to creative thinking. New York: Atherton Press, 1962.

Burt, Cyril. The sign pattern of factor-matrices. The British Journal of Statistical Psychology, Vol. VII (1954), 15-28.

Campbell, Donald T. & Fiske, Donald W. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, LVI (March, 1959), 81-105.

Campbell, Donald T. Recommendations for APA test standards regarding construct, trait, or discriminant validity. The American Psychologist, 1960, 546-553.

Canisia, Sister M. Mathematical ability as related to reasoning and use of symbols. Educational and Psychological Measurement, 1962, 22, 105-127.

Carlton, Lucia V. An analysis of the educational concepts of fourteen outstanding mathematicians, 1790-1940. Unpublished doctoral dissertation, Northwestern University, 1959.

Cattell, Raymond B., & Butcher, H. J. The prediction of achievement and creativity. Indianapolis/New York: The Bobbs-Merrill Company, Inc., 1968.

Christofferson, H. C. Creative thinking in mathematics. The Mathematics Teacher, LI (November, 1958), 535-540.

Cronbach, L. J. The learning by discovery hypothesis. In Lee S. Shulman & Evan R. Keislar (Eds.), Learning by discovery: A critical appraisal. Chicago: Rand McNally & Company, 1966.

Cronbach, L. E., & Meehl, Paul E. Construct validity in psychological tests. Psychological Bulletin, Vol. LII (1955), 281-301.

Cronach, L. J. Essentials of psychological testing. New York: Harper & Row, 1970.

Davis, Robert B. The next few years. The Arithmetic Teacher, XIII (May, 1966), 335-362.

Davis, Robert B. Discovery in the teaching of mathematics. In Lee S. Shulman & Evan R. Keislar (Eds.), Learning by discovery; A critical appraisal. Chicago: Rand McNally & Company, 1966.

De Bono, Edward. Information processing and new ideas -- lateral and vertical thinking. The Journal of Creative Behavior, Vol. III (1969), 150-171.

De Cecco, John P. The psychology of learning and instruction: Educational psychology. Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1968.

Einstein, Albert. (A letter to Hadarmard). In Jacques Hadarmard, The psychology of invention in the mathematical field. Dover Publications, Inc., 1945, p. 142.

Eisner, Elliot W. Think with me about creativity. New York: Owen Publishing Company, 1964.

Evans, Edward E. Measuring the ability of students to respond in creative mathematical situations at the late elementary and junior high school level. Unpublished Ph.D. dissertation, University of Michigan, 1964.

Falkoff, A. D., & Iverson, K. E. APL/360: User's manual. New York: International Business Machines Corporation, 1968.

Ferguson, George A. Statistical analysis in psychology and education. New York: McGraw-Hill Book Company, 1966.

Getzels, Jack & Madaus, George F. Creativity. Encyclopedia of Educational Research, 196 .

Ghiselin, B. The creative process and its relation to the identification of creative talent. In C. W. Taylor (Ed.), The 1955 University of Utah Research Conference on the identification of creative talent. Salt Lake City: University of Utah Press, 1956.

Ghiselin, B. (Ed.) The creative process. The New American Library, 1952.

Guilford, J. P. Three faces of intellect. American Psychologist, XIV (August, 1959), 469-479.

Guilford, J. P. Traits of creativity. In Harold H. Anderson (Ed.), Creativity and its cultivation. New York & Evanston: Harper & Row, 1959.

Guilford, J. P. A psychometric approach to creativity. In Harold H. Anderson (Ed.), Creativity in childhood and adolescence. California: Science and Behavior Books, Inc., 1965.

Guilford, J. P. Creativity. American Psychologist, V (September, 1950), 444-454.

Guilford, J. P. Creativity: Its measurement and development. In Sydney J. Parnes and Harold F. Harding (Eds.), A source book for creative thinking. New York: Charles Scribner's Sons, 1962.

Guilford, J. P. Some theoretical views of creativity. In Harry Helson & William Bevan (Eds.), Contemporary approaches to psychology. Princeton, N. J.: D. Van Nostrand Company, Inc., 1967b.

Guilford, J. P. The nature of human intelligence. New York: McGraw-Hill, 1967.

Guilford, J. P. Sixteen divergent-production abilities at the ninth-grade level. Multivariate Behavioral Research, (January, 1966), 43-67.

Guilford, J. P. Some new look at the nature of creative processes. In Norman Frederiksen and Harold Gulliksen (Eds.), Contributions to mathematical psychology. New York: Holt, Rinehart & Winston, Inc., 1964.

Guilford, J. P. Psychometric methods. New York: McGraw-Hill, 1954.

Guilford, J. P., et al. Predicting achievement in ninth-grade mathematics from measures of intellectual-aptitude factors. Educational and Psychological Measurement. XXV (1965), 659-681.

Gulliksen, H. Theory of mental tests. New York: Wiley, 1950.

Guttman, Louis. Best possible: Systematic estimate of communalities. Psychometrica, Vol. XXI (1956), 273-285.

Guttman, Louis. A new approach to factor analysis: The radex. In Paul F. Lazarfeld (Ed.), Mathematical thinking in the social sciences. New York: Columbia University Press, 1954, 258-348.

Hadarmard, Jacques. The psychology of invention in the mathematical field. Dover Publications, Inc., 1945.

Harding, Harold F. The need for a more creative trend in American education. In Sydney J. Parnes & Harold F. Harding (Eds.), A source book for creative thinking. New York: Charles Scribner's Sons, 1962.

Harker, Frederick J. Creative possibilities for a consideration of creativity. In Harold H. Anderson (Ed.), Creativity in childhood and adolescence. California: Science and Behavior Books, Inc., 1965.

Harman, Harry H. Modern factor analysis. Chicago: The University of Chicago Press, 1960.

Henle, Mary The birth and death of ideas. In Glen Terrell & Michael Wertheimer (Eds.), Contemporary approaches to creative thinking. New York: Atherton Press, 1962.

Herstein, I. N. Topics in algebra. New York: Blaisdell Publishing Company, 1964.

Hilgard, Ernest R. (Ed.) Theories of learning and instruction. The Sixty-third Yearbook of the National Society for the study of Education. The University of Chicago Press, 1964.

Hirschi, L. Edwin Encouraging creativity in the mathematics classroom. The Mathematics Teacher, LVI (February, 1963), 79-83.

Hohn, F. E. Teaching creativity in mathematics. The Arithmetic Teacher, VIII (March, 1961), 102-106.

Hunka, S. Research and information report. Introduction to APL 360/67 programming. Edmonton: Division of Educational Research Services, Faculty of Education, The University of Alberta, November, 1967.

Husen, Torsten International study of achievement in mathematics. New York: John Wiley and Sons, 1967.

Jackson, D. J. & Messick, Samuel (Eds.) Problems in human assessment. New York: McGraw-Hill, 1967.

Jackson, H. C. Creative thinking and discovery. The Arithmetic Teacher, VIII (March, 1961), 107-111.

Jackson, P. W. and Messick, Samuel The person, the product, and the response: Conceptual problems in the assessment of creativity. Journal of Personality, XXXIII, 1965, 309-329.

Johnson, R. J. The mathematizing mode. Unpublished master's thesis, University of Alberta, 1968.

Khinchin, A. Ya. The teaching of mathematics. New York: American Elseview Publishing Company, Inc., 1968.

Kneller, G. F. The art and science of creativity. New York: Holt, Rinehart & Winston, Inc., 1965.

Koestler, A. The act of creation. London: Hutchinson, 1964.

Loevinger, Jane. Objective tests as instruments of psychological theory. In Douglas N. Jackson & Samuel Messick (Eds.), Problems in human assessment. New York: McGraw-Hill, 1967.

Lowenfeld, Nikfor. Creativity: Education's stepchild. In Sydney J. Parnes & Harold H. Harding (Eds.), A source book for creative thinking. New York: Charles Scribner's Sons, 1962.

Maguire, T. O., & Hazlett, C. B. Reliability for the researcher. Alberta Journal of Educational Research, Vol. XV, 1969, 117-125.

May, Frank B. & Metcalf, Alan W. A factor-analytic study of spontaneous-flexibility measures. Educational and Psychological Measurement, Vol. XXV, 1965, 1039-1049.

Mead, Margaret. Creativity in cross-cultural perspective. In Harold H. Anderson (Ed.), Creativity and its cultivation. New York & Evanston: Harper & Row, 1959.

Meconi, L. J. The mathematically gifted student and discovery learning. The Mathematics Teacher, LX, December, 1967, 862-865.

Michel, L. Allgemeine grundlagen psychometrischer tests. In Prof. Dr. R. Heiss (ed) Psychologische Diagnostik, Gottingen: Verlag fur Psychologie, Dr. C. J. Hogrefe, 1964.

Moise, E. E. Elementary geometry from an advanced standpoint. Reading, Mass.: Addison-Wesley Publishing Co. Inc., 1963.

Mordell, L. J. Reflections of a mathematician. Montreal: Canadian Mathematical Congress, 1959.

Morrison, D. F. Multivariate statistical methods. New York: McGraw-Hill, 1967.

Osborn, A. F. Applied imagination: Principles and procedures of creative thinking. New York: Charles Scribner's Sons, 1953.

Osborne, Alex F. Developments in creative education. In Sydney J. Parnes & Harold F. Harding (Eds.), A source book for creative thinking. New York: Scribner's Sons, 1962.

Overall, J. E. Reliability of composite ratings. Educational and Psychological Measurement, XXV, 1965, 1011-1022.

Parnes, Sydney J. & Harding, Harold F. A source book for creative thinking. New York: Scribner's Sons, 1962.

Parnes, Sydney J. The nurture of creative talent. Music Educators Journal, June-July, 1965, 32-3;92-5.

Patrick, Catherine. What is creative thinking? New York: Philosophical Library, 1955.

Poincare, Henri. Mathematical creation. In B. Ghiselin (Ed.), The creative process. The New American Library, 1952.

Polya, George. On teaching problem solving. In The role of axiomatics and problem solving in mathematics. The Conference Board of the Mathematical Sciences. Washington, D.C.: Ginn and Company, 1966.

Prouse, H. L. Creativity in school mathematics. The Mathematics Teacher, LX, December, 1967, 876-879.

Rogers, Carl F. Towards a theory of creativity. In Sydney J. Parnes & Harold F. Harding (Eds.), A source book for creative thinking. New York: Charles Scribner's Sons, 1962.

Scheerer, Martin. Problem solving. Scientific American, April, 1963, 2-9.

Sigurdson, S. E. The mathematising project. Unpublished report prepared for teachers who participated in a Maths 20 project. Edmonton: Department of Secondary Education, University of Alberta, September, 1968.

Stein, Morris I. Creativity as an intra- and inter-personal process. In Sydney J. Parnes and Harold F. Harding (Eds.), A source book for creative thinking. New York: Charles Scribner's Sons, 1962.

Stein, Morris I., & Heinze, Shirley J. Creativity and the individual. The Graduate School of Business, The University of Chicago, 1960.

Steiner, Hans-George. Examples of exercies in mathematization on the secondary school level. Educational Studies in Mathematics, I, January, 1967, 289-299.

Taba, Hilda. Curriculum development: Theory and practice. New York: Harcourt, Brace & World, 1962.

Taylor-Pearce, J. Modupe. A study of the relative effectiveness of two teaching methods with respect to (divergent thinking): Creativity in mathematics at the grade eleven level. Unpublished master's thesis, University of Alberta, 1969.

Thorndike, R. L., et al. Reliability. In Douglas N. Jackson & Samuel Messick (Eds.), Problems of human assessment. New York: McGraw-Hill Book Company, 1967.

Thurstone, L. L. Multiple factor analysis. Chicago: The University of Chicago Press, 1947.

Torrance, E. Paul. Developing creative thinking through school experiences. In Sydney J. Parnes & Harold F. Harding (Eds.), A source book for creative thinking. New York: Charles Scribner's Sons, 1962.

Torrance, E. Paul. Rewarding creative behavior. Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1965.

Torrance, E. Paul. Guiding creative talent. Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1963.

Torrance, E. Paul. Examples and rationales of test tasks for assessing creative abilities. Journal of Creative Behavior, II, (Summer, 1968), 165-177.

Torrance, E. Paul. Role of evaluation in creative thinking. University of Minnesota: Cooperative Research Project, No. 125, 1964.

Wallas, Graham. The art of thought. London: C. A. Watts and Co., Ltd. (1926), 1945.

Wallen, C. J. Stressing the creative aspects of mathematics in teaching the gifted child. The Mathematics Teacher LV, April, 1962, 243-48.

Wertheimer, Michael. Productive thinking. New York: Harper & Brothers, 1959.

Whiting, Charles C. Creative thinking. New York: Reinhold, 1958.

Wilson, Robert C., Guilford, J. P., & Christensen, Paul R. A factor-analytic study of creative thinking abilities. Psychometrica, XIX, December, 1954.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill Book Company, 1962.

Worthen, Blaine R. A comparison of discovery and expository sequencing in elementary mathematics instruction. In Joseph M. Scandura (Ed.), Research in mathematics education. Washington, D.C.: The National Council of Teachers of Mathematics Inc., 1967.

Youtz, Richard P. Psychological foundations of applied imagination. In Sydney J. Parnes and Harold F. Harding (Eds.), A source book for creative thinking. New York: Scribner's Sons, 1962.

APPENDIX A

SOME TABLES

TABLE I

KEY TO
LETTER SYMBOLS DENOTING EXPRESSIONS FREQUENTLY USED IN CONNECTION WITH DIVERGENT PRODUCTION

Symbol	Meaning
C	Classes
DMaI	Divergent Production of Mathematical Implications
DMaR	Divergent Production of Mathematical Realtions
DMaS	Divergent Production of Mathematical Systems
DMaT	Divergent Production of Mathematical Transformations
DMaU	Divergent Production of Mathematical Units
DMaI(F)	Facility in DMaI
DMaS(F)	Facility in DMaS
DMaT(F)	Facility in DMaS
DMaI(F)	Facility in DMaI
DMaS(V)	Variety in DMaS
DMaT(V)	Variety in DMaT
DMaU(V)	Variety in DMaU
DMaI(N)	Novelty in DMaI
DMaS(V)	Novelty in DMaS
DMaT(V)	Novelty in DMaT
DMaU(V)	Novelty in DMaU
DP	Divergent Production
F	Facility
FMI	Facility in Mathematical Implications
FMS	Facility in Mathematical Systems

TABLE 1 (CONT'D)

KEY TO
LETTER SYMBOLS DENOTING EXPRESSIONS FREQUENTLY USED IN CONNECTION WITH DIVERGENT PRODUCTION

Symbol	Meaning
FMT	Facility in Mathematical Transformations
FMU	Facility in Mathematical Units
I	Implications
N	Novelty
NMI	Novelty in Mathematical Implications
NMS	Novelty in Mathematical Systems
NMT	Novelty in Mathematical Systems
NMU	Novelty in Mathematical Units
S	Systems
T	Transformations
V	Variety
VMI	Variety in Mathematical Implications
VMS	Variety in Mathematical Systems
VMT	Variety in Mathematical Transformations
VMU	Variety in Mathematical Units

TABLE 2

CLASSIFICATION OF STUDENT RESPONSES TO "STABLE" UNITS TESTS 1 AND I IN TERMS OF CLASSES OF INVENTIVE RESPONSES

MAIN SAMPLE 1

Student	Number of responses in Class 00	Class 10	Class 15	Class 20	Class 25	Class 30[1]
1	6	4	4	1	0	0
2	2	5	9	4	2	1
3	10	6	3	1	0	0
4	6	2	3	0	0	0
5	4	2	1	1	1	1
6	1	4	1	1	0	0
7	3	3	4	0	0	0
8	3	3	1	1	0	0
9	2	1	0	0	0	0
10	5	1	4	1	0	0
11	3	3	3	0	0	0
12	7	1	1	1	0	0
13	3	2	1	0	0	0
14	2	1	1	0	1	0
15	8	5	5	0	0	0
16	2	4	3	5	0	0
17	3	2	2	1	0	0
18	0	6	2	0	0	0
19	6	3	1	0	0	0
20	6	4	2	2	0	0
21	7	1	0	1	0	0
22	5	1	1	1	0	0
23	5	3	3	0	0	0
24	6	3	8	3	1	1
25	9	10	4	1	0	0

[1]Class 30 was the highest class containing responses.

TABLE 2 (CONTINUED)

CLASSIFICATION OF STUDENT RESPONSES TO "STABLE" UNITS TESTS 1 and I
IN TERMS OF CLASSES OF INVENTIVE RESPONSES

MAIN SAMPLE 1

Student	Number of responses in Class 00	Class 10	Class 15	Class 20	Class 25	Class 30
26	5	2	3	1	0	0
27	5	2	2	1	0	0
28	3	1	2	0	0	0
29	3	0	1	0	0	0
30	15	2	0	0	0	0
31	4	2	4	1	0	0
32	4	2	2	0	0	0
33	3	1	1	1	0	0
34	3	2	3	4	1	1
35	7	4	6	2	1	0
36	7	4	4	4	1	0
37	4	6	2	6	2	0
38	6	3	7	2	2	1
39	3	2	3	1	0	0
40	6	2	1	1	0	0

TABLE 2 (CONTINUED)

CLASSIFICATION OF STUDENT RESPONSES TO "STABLE" UNITS TESTS 1 and I IN TERMS OF CLASSES OF INVENTIVE RESPONSES

MAIN SAMPLE 2

Student	Class 00	Number of responses in Class 10	Class 15	Class 20	Class 25	Class 30
1	5	4	1	0	0	0
2	7	1	1	1	0	0
3	4	3	0	1	0	0
4	3	2	1	1	0	0
5	7	2	1	0	0	0
6	4	3	3	1	0	0
7	2	4	3	0	0	0
8	10	4	2	0	1	0
9	5	1	1	0	0	0
10	6	3	2	1	0	0
11	3	1	3	1	1	0
12	0	0	0	0	0	0
13	3	3	2	0	0	0
14	1	0	0	0	0	0
15	7	2	0	0	0	0
16	2	1	0	1	0	0
17	2	3	2	0	0	0
18	4	4	4	0	0	0
19	10	5	0	0	0	0
20	3	0	4	1	0	0
21	1	2	2	0	0	0
22	0	0	2	1	0	0
23	4	2	0	0	0	0
24	3	3	2	1	0	0
25	2	0	0	0	0	0
26	5	4	1	1	0	0
27	4	1	2	0	0	0
28	1	1	1	1	0	0
29	1	2	3	2	0	1

TABLE 2 (CONTINUED)

CLASSIFICATION OF STUDENT RESPONSES TO "STABLE" UNITS TESTS 1 and I IN TERMS OF CLASSES OF INVENTIVE RESPONSES

MAIN SAMPLE 2

Student	Class 00	Number of responses in Class 10	Class 15	Class 20	Class 25	Class 30
30	3	0	0	1	0	0
31	7	2	1	0	0	0
32	3	4	2	1	0	0
33	3	1	2	1	0	0
34	5	1	5	0	0	0
35	8	3	4	0	0	0
36	3	2	0	0	0	0
37	6	2	5	1	0	0
38	5	1	0	0	0	0
39	9	2	3	1	0	0
40	6	9	1	1	0	0
41	3	2	0	1	0	0
42	6	1	0	0	0	0
43	10	5	4	4	0	0
44	4	2	1	1	0	0
45	8	5	4	5	0	0
46	6	1	4	2	0	0
47	12	4	6	5	0	0
48	2	1	2	0	0	0
49	2	3	2	0	0	0
50	2	1	3	1	0	0
51	2	0	1	2	0	0
52	2	1	0	0	0	0
53	2	1	0	0	0	0
54	4	2	0	0	0	0
55	4	7	1	0	0	0
56	6	4	3	2	0	0
57	5	1	2	1	0	0
58	6	1	0	0	1	0

TABLE 2 (CONTINUED)

CLASSIFICATION OF STUDENT RESPONSES TO "STABLE" UNITS TESTS 1 and I IN TERMS OF CLASSES OF INVENTIVE RESPONSES

MAIN SAMPLE 2

	Number of responses in					
Student	Class 00	Class 10	Class 15	Class 20	Class 25	Class 30
59	4	2	1	0	0	0
60	1	0	2	1	0	0
61	5	2	5	1	0	0
62	4	0	0	0	0	0

TABLE 3

CLASSIFICATION OF STUDENT RESPONSES TO "STABLE" CLASSES
TESTS 2 AND II

MAIN SAMPLE 1

Student	Class 00	Class 10	Class 15	Class 20	Class 25
1	0	2	0	3	2
2	0	2	0	0	0
3	0	1	0	2	0
4	0	2	0	2	0
5	0	1	0	2	1
6	0	2	0	3	0
7	0	2	0	1	0
8	0	2	0	0	0
9	0	2	0	1	0
10	0	1	0	1	0
11	0	2	0	0	0
12	0	2	0	1	0
13	0	2	0	1	0
14	0	2	0	2	0
15	0	2	0	1	0
16	0	2	0	3	1
17	0	2	0	0	0
18	0	2	0	0	0
19	0	2	0	0	0
20	0	2	0	2	1
21	0	2	0	0	0
22	0	2	0	1	0
23	0	2	0	3	0
24	0	2	0	2	0
25	0	2	0	2	0
26	0	2	0	0	0
27	0	2	0	0	0

TABLE 5 (CONTINUED)

CLASSIFICATION OF STUDENT RESPONSES TO "STABLE" CLASSES
TESTS 2 AND II

MAIN SAMPLE 1

Student	Class 00	Class 10	Class 15	Class 20	Class 25
28	0	2	0	0	0
29	0	1	0	0	0
30	0	2	0	0	0
31	0	2	0	0	0
32	0	2	0	0	0
33	0	2	0	0	0
34	0	2	0	2	0
35	0	2	0	3	1
36	0	2	0	2	1
37	0	2	0	2	0
38	0	2	0	3	0
39	0	1	0	0	0
40	0	2	0	2	1

TABLE 3 (CONTINUED)

CLASSIFICATION OF STUDENT RESPONSES TO "STABLE" CLASSES
TESTS 2 AND II

MAIN SAMPLE 2

Student	Class 00	Class 10	Class 15	Class 20	Class 25
1	0	2	0	1	0
2	0	2	0	3	0
3	0	2	0	3	0
4	0	2	0	2	1
5	0	2	0	3	1
6	0	2	0	0	0
7	0	2	0	3	0
8	0	2	0	0	0
9	0	1	0	1	0
10	0	0	0	0	0
11	0	2	0	2	0
12	0	2	0	2	1
13	0	0	0	0	0
14	0	2	0	0	0
15	0	1	0	1	1
16	0	2	0	2	1
17	0	2	0	1	0
18	0	2	0	0	0
19	0	1	0	1	2
20	0	1	0	2	0
21	0	1	0	0	0
22	0	0	0	0	0
23	0	0	0	0	0
24	0	2	0	1	0
25	0	2	0	2	0
26	0	1	0	0	0
27	0	2	0	3	1
28	0	2	0	0	0

TABLE 3 (CONTINUED)

CLASSIFICATION OF STUDENT RESPONSES TO "STABLE" CLASSES
TESTS 2 AND II

MAIN SAMPLE 2

Student	Class 00	Class 10	Class 15	Class 20	Class 25
29	0	2	0	1	0
30	0	0	0	0	0
31	0	2	0	2	0
32	0	1	0	1	0
33	0	2	0	2	0
34	0	1	0	0	0
35	0	2	0	0	0
36	0	1	0	0	0
37	0	2	0	0	0
38	0	2	0	0	0
39	0	2	0	2	0
40	0	2	0	2	0
41	0	2	0	3	0
42	0	1	0	1	0
43	0	2	0	1	0
44	0	2	0	3	0
45	0	1	0	1	0
46	0	2	0	3	1
47	0	1	0	2	0
48	0	2	0	0	0
49	0	1	0	1	0
50	0	1	0	1	0
51	0	2	0	0	0
52	0	1	0	1	0
53	0	0	0	0	0
54	0	1	0	0	0
55	0	2	0	0	0
56	0	2	0	1	0
57	0	2	0	0	0

TABLE 3 (CONTINUED)

CLASSIFICATION OF STUDENT RESPONSES TO "STABLE" CLASSES
TESTS 2 AND II

MAIN SAMPLE 2

Student	Class 00	Class 10	Class 15	Class 20	Class 25
58	0	1	0	0	0
59	0	1	0	0	0
60	0	2	0	2	0
61	0	2	0	1	0
62	0	2	0	1	0

APPENDIX B

REPORT OF PILOT STUDY

REPORT ON PRELIMINARY STUDY

The preliminary study was conducted at a high school in Edmonton. The tests were administered by five subject teachers to their classes. The investigator was given the opportunity to hold two meetings with the subject teachers in which he explained the aims of the study and the purpose of the tests. The test administrators were later given a summary of the administrative procedures in writing.

The tests were administered over a period of four consecutive days in January, 1970. They were arranged in four batteries, each battery containing one test for each ability, in a single booklet. Each of the four booklets had seven pages, and each page contained one problem. Ten minutes were allowed for each of the DP tests, and twenty minutes for the CPS tests. Two hundred and fifteen students attempted at least one of the batteries, but only sixty-four attempted all four batteries. These 64 constitute the sample of the pilot study.

The teachers, except one, indicated that the students in general were not motivated to participate in the experiment. In one case, a teacher informed the investigator, after the administration of the second battery, that her students were dissatisfied with the testing, and wished to meet the investigator. The investigator discussed with the students, the next day, explained the purpose of the tests, the potentialities of the experimentation, and the relevance of invention in mathematics, and had a pleasant time of discussion and questioning. The students showed considerable interest during the discussion, and nearly all volunteered to resume the experiment the next day. The teacher was evidently highly impressed with the discussion, the participation of the students, and particularly with

the number that volunteered to continue. More than one teacher advised the investigator to have a talk with the students in a future experiment prior to their embarking on the experiment. The one teacher who indicated that his students were motivated -- and this was borne out by the attendance in his class - explained that he spoke to the students about the purpose of the tests, and encouraged them to participate in the testing actively.

It was observed in marking the papers of the sixty four students that many of them did not write down anything for some of the tests within a battery. In some cases, all the students in a class did not attempt a particular test. In other cases, matters obviously not relevant to the tests were written down, and in others, what was written down was simply not appropriate. Scores obtained under such conditions were clearly of doubtful validity. However, the investigator computed certain psychometric relationships among the tests, to assist in making decisions on the final selection.

The responses were marked for facility in production. One mark was given for each appropriate response made by a student. A response was appropriate if it satisfied the basic requirements of the test situation, even if there were minor arithmetic errors in the response.

A divergent productivity ratio (P) was calculated for each test. P was defined as the ratio of the number of students who produced more than one response, to the total number of students in the sample. Reliability estimates on raw scores for the mean of all the tests in a product category, using analysis of variance techniques (Winer, 1962, p. 124-132.) Where a student failed to score in a test,

he was awarded zero. The reliability estimates were moderate for the test in Units (.69) and Classes (.67), but low in Relations (.20), Systems (.16), Transformations (.14) and Implications (.19). The productivity ratios were found to be as follows:

Category	Test No.	P
Units	I	.72
	1	.72
	IB	.42
	1B	.33
Classes	II	.11
	2	.33
	IIB	.16
	2B	.22
	IIB	.16
	2B	.22
Relations	III	.44
	3	.55
	IIIB	.28
	3B	.47
Systems	IV	.02
	4	.30
	IVB	0
	4B	.50

Category	Test No.	P
Trans.	V	0
	5	.06
	VB	.44
	5B	.25
Imp.	VI	.60
	6	.36
	VIB	.03
	6B	.19

Final Selection--Rationale

The final selection of tests were made with the following considerations in mind.

1. That each test selected should have been shown to lead to divergent production by having a positive productivity ratio.
2. That each test selected should have in the judgement of the investigator resulted in inventive productions.
3. That when a test is selected, at least one more test should be selected in the same product category.
4. That the total number of tests selected should be small.

Discussion

The first principle of selection was based on the consideration that the tests were tests of divergent production, and hence should lead to the production of more than one appropriate response. The second principle was based on the purpose of the study, which was to construct tests of inventiveness in mathematics. Thus the validity of the tests depended largely on their being capable of consistently

evoking inventive responses. The third principle was due to the fact that at least two tests are needed to make certain important decisions on the validity of the tests in determining constructs. The fourth principle was due to the long hours needed to mark divergent productions. It was felt that meaningful information would be obtained from a restricted number of tests.

It was not considered essential to the problem of the study to take all Guilford's product categories into account in testing in each of the two samples. The primary interest was in the tests and what they measured. The Relations category was dropped from the study. It was felt that the production of Classes implied an appreciation of certain relationships of similarity and differences. Ten DP tests were selected from the remaining five product categories for testing in one sample, and another ten were selected from three categories for testing in the other. Six of the DP tests were common to the two samples. The use of two main samples and different tests afforded opportunities for comparisons leading to observations of the stability of constructs, which a single sample would not afford.

APPENDIX C

STANDARDIZED RESPONSES

FROM

SOME OF THE TESTS

TEST 1

Write down as many mathematically true statements as you can about an Epudom in the sense defined below:

An Epudom is an integer divisible by 35.

I.D. No. Responses

1. Every Epudom is divisible by -35.

2. Every Epudom is divisible by 35.

3. Every Epudom is divisible by 7.

4. Every Epudom is divisible by -7.

5. Every Epudom is divisible by 1.

7. Every Epudom is divisible by -1.

8. Every Epudom is divisible by 5.

9. Every Epudom is divisible by -5.

10. Epudoms are not primes.

11. If a line is drawn through points representing all Epudoms ....it will be straight.

12. The slope of the 'Epudom line' is 35.

13. The set of all Epudoms is infinite.

14. Epudoms can be positive or negative.

15. Any Epudom (E) can be found by $35x = E$, where $x \in I$.

16. Any Epudom can divide only other Epudoms.

17. All Epudoms are integers, rational numbers, real numbers.

18. There are as many Epudoms on one side of zero as on the other.

19. Each negative Epudom is 35 times a negative number.

20. Each positive Epudom is 35 times a positive number.

21. There is an additive identity (0) in the set of Epudoms.

I.D. No. Responses

22. An Epudom raised to any power is an Epudom. i.e., $E^n \in E$.

23. The set of Epudoms form a <u>number system</u>.
(Note: "A number system is a set of entities on which are defined two binary operations called addition and multiplication, such that the set is closed under each operation, and each operation is commutative and associative, and multiplication is distributive over addition.)

24. Epudoms are closed with respect to addition and multiplication.

25. Epudoms are commutative and associative with respect to addition and multiplication.

28. There is no multiplicative inverse property in the set of Epudoms.

31. Epudoms do not form a <u>field</u>.

32. Epudoms may be even or odd.

34. Epudoms will always end in 0 or 5.

35. Epudoms will increase or decrease in 35 unit intervals.

36. Epudoms can be as large as you can get.

37. Epudoms can be as small as you can get.

37b. Epudoms are infinite in both directions.

38. Examples of Epudoms. (70 is an E).

41. The sum of the first and last digits of an Epudom never exceed 14.

42. Between any two numbers, 100 in difference, there must be a minimum of two and a maximum of three Epudoms.

44. By adding 1 to an odd Epudom, it becomes even and vice versa.

45. Epudom + 4 = 4 + Epudom.

46. Epudom x 1 = Epudom.

46b. An Epudom is divisible by all factors of 35.

47. All odd Epudoms end in 5.

52. In the statement 35x = Epudom, if x is odd, then Epudom is odd. If x is even, then Epudom is even.

I.D. No. Responses

53. If $ab = E$, then E is positive whenever a and b are both negative or both positive. E is negative otherwise.

54. Epudoms could form a new number system with 35 being equivalent to 1 in the arabic system. All numbers are divisible by it. Only whole numbers are allowed in this system.

55. Epudoms are closed with respect to subtraction.

57. Ordered pairs may be set to represent the situation: e.g., (-1,-35); (1,35).

60. The slope of a line parallel to the Epudom line is 35.

61 The slope of the line perpendicular to the Epudom line is $-\frac{1}{35}$.

62. The graph of the Epudoms is a series of dots because the set of numbers involved is the integers.

64. If Epudom is odd, then Epudom/35 is odd.

65. If Epudom is even, Epudom/35 is even.

66. If Epudom is negative, E/35 is negative.

67. If E is positive, then E/35 is positive.

71. If E is exactly divisible by 2, it is an even number.

72. If an E is evenly divisible by an even number, it is an even number.

74. The Epudom for which 35 is multiplied by an even number ends in 0 -- by odd it ends in 5.

75. The set of Epudoms obeys the distributive property.

76. $\log_{35}E = 1 + \log_{35}E/35$.

77. 35 and 70 are the only two Epudoms between 1 and 100.

79. If Epudom = A, then $(A+5) - (2A) = (-A) + 5$.

80. Every even Epudom is divisible by 10.

81. If $a \in \{\text{Epudom}\}$, $a/35 \times 35/a = 1$

82. a/0 is undefined.

I.D. No. Responses

83. a = a.

84. If a/35 is greater than 0, then a is greater than 0.

85. If a/35 is less than 0, then a is less than 0.

86. If a/35 equals 0, then a equals 0.

96. For every E = 35x, where x is an integer, $|E|!$ will always be greater than either E or $|E|$.

102. The set of Epudoms is equivalent to the set of integers.

103. Since the set of Epudoms has closure, associative, inverse, identity properties under addition, it is a group with respect to addition.

104. Since it is commutative, it is an abelian group.

105. The integers which result when the set of Epudoms is divided by 35 make up the set of integers.

106. An integer divisible by 70 is also an Epudom.

107. An Epudom x 2 will always end in 0.

108. Epudoms are not closed with respect to division.

109. $E + 35 \in \{E\}$.

110. $E - 35 \in \{E\}$.

111. $E \times 35 \in \{E\}$.

113. There are as many odd as there even Epudoms. (Also as many odd as there are <u>all</u> Epudoms).

114. $E \times 2 \in \{E\}$.

115. E x any integer is also divisible by 35.

118. The Epudom can have an infinite number of digits.

119. The smallest positive E is 35.

120. The largest negative E is -35.

121. There is no highest or lowest E.

I.D. No.	Responses
122.	Not all numbers divisible by 7 are Epudoms
123.	Not all numbers divisible by 5 are Epudoms.
124.	E is divisible by itself.
125.	This is a function: E = 35x.
127.	As each Epudom increases, the sum of the numbers (digits) decreases by 1, up to 4. Note: 35 x 1 = 35 S(digits) = 8 35 x 2 = 70 S(digits) = 7 35 x 3 = 105 S(digits) = 6
128.	Any E can be factored into 5 x 7 x (E/35).
129.	E/35 x 7 = E/5. / E/35 x 5 = E/7.
130.	Every even E is divisible by 70.
131.	1 1/2 is not an E.

TEST I

Write down as many mathematically true statements as you can about a Shola in the sense defined below:

A Shola is an odd integer divisible by 39.

I.D. Response

4. Sholas are divisible by 13(or -13).

4b. Sholas are divisible by 39(or -39).

2. Sholas are divisible by 1(or -1).

3sp. The sum of digits in a shola is divisible by 3.

3. Sholas are divisible by 3(or -3).

5. S = 39(2n+1), where n is an integer, and S denotes Sholas.

7. A Shola cannot be evenly divided by 2.

8. All factors of 39 are factors of a Shola. (Each Shola is divisible by all factors of 39).

9. A Shola is a composite number (not prime).

10. Shola x even integer = even integer.

11. Shola x odd integer = odd integer.

12. If a Shola is expressed as 39x, x is always odd.

13. Shola/3 is an odd integral number.

14. Shola/13 is an odd integral number.

15. $(\text{Shola})^x$, where x is a positive integer, is an odd number.

16. There is an infinite number of Sholas.

17. Sholas are not closed with respect to addition.

18. Sholas are closed with respect to multiplication.

19. Sholas do not form a number system.

(Investigator's Note: A number system is defined in students' textbook as a set of entities, on which two binary operations (addition and multiplication) are defined, such that the set is closed under each operation, and each operation is commutative and associative, and multiplication is distributive over addition).

I.D. Response

20. Sholas are commutative with respect to addition and multiplication.

22. Sholas are associative with respect to addition and multiplication.

23. Sholas are of the form (2n+1), where n is an integer.

24. A Shola can be as big as you wish.

25. A Shola can be located on the number line.

26. A Shola will have two or more digits.

27. 39 is the only Shola below 100.

28. If Shola = 39x, then Shola is divisible by x. Shola is divisible by Shola/39.

29. Sholas are positive and negative.

31. Sholas are real numbers.

33. 39 multiplied by an even number cannot be a Shola.

34. Specific examples of Sholas: 39,-39,117,-117.

35. Add or subtract 78 to a Shola to obtain the nearest Shola.

36. If a Shola is negative, then Shola/39 will be negative.

37. 1 x Shola = Shola.

38. 0 x Shola = 0

39. Shola/0 is undefined.

40. Shola + 1 $\neq$ Shola

41. Shola $\pm$ an integer is an integer.

43. Subtract or add a Shola to a Shola, and the result will be divisible by 2.

46. The factors of a Shola are always odd.

47. Shola can be a square, cube, etc.

48. As Sholas increase positively from 39, the last digits decrease: 9,7,5,3,1,9 ...... It then starts again at 9.

I.D. Response

50. Let S_1, S_2, S_3, be elements in the set of Sholas. Then $S_1 + S_2 + S_3$ is a Shola.

51. In general, any sum of an even number of Sholas is not a Shola, while the sum of an odd number of Sholas is a Shola.

52. A one-one correspondence ben be set between the set of all Sholas and the set of integers.

56. There is no additive identity element in the set of Sholas.

57. If S is a Shola, then $S \geq 39$, or $S \leq -39$.

59. There is no multiplicative inverse in the set of Sholas.

60. There is no multiplicative identity element in the set of Sholas.

67. If x is a Shola, x may be a unique real number. If x is negative, then x will not be a real number, but may be defined using i such that $i^2 = -1$.

63. The absolute value of a Shola is always greater than or equal to 39.

69. A Shola plus an odd integer is always even.

70. If x is a Shola, then x^2 is a Shola.

71. If x is a Shola, then $(x+1)^2$ is an even integer.

72. Shola + Shola = even integer.

73. The sum of a Shola and an even integer is an odd integer.

74. A Shola multiplied by another Shola is an odd integer.

78. Zero is not a Shola.

81. Sholas are not closed with respect to division.

82. The graph of S = 39x, where x is an odd integer, is linear.

83. Some ordered pairs of 82 are: (1,39), (-1,-39), (3,117).

84. The relation S = 39x, is a function.

87. Shola/39 is an odd factor of Shola.

89. Every Shola is divisible by itself.

I.D.	Response
91.	Examples of non-Sholas. e.g., 78. Divisible by 39 and yet not a Shola.
93.	Log Shola = log 39 + log x, where x is an odd integer.
94.	13 x (S/39) = S/3.
95.	Any Shola can be factored into 3 x 13 x n, where n is any odd integer, positive or negative.
98.	The Shola graph lies in the first and third quadrants.
99.	$\{x: \log_{39}x$ is an odd integer$\}$ Any number satisfying this condition is a Shola. N.B. All Sholas will <u>not</u> satisfy the condition.
101.	If the product of a Shola and a real number results in a Shola, that real number is an odd integer.
102.	39 is the only Shola between 1 and 100.
105.	The number obtained when a Shola is multiplied by any integer, is divisible by 39.
106.	The sum of two Sholas is divisible by 39.
108.	A Shola may have an infinite number of digits.
109.	Shola/39 can equal a Shola.
111.	There are 13 positive Sholas less than 1000.
113.	There is no greatest or smallest Shola.
114.	Not all odd numbers are Sholas.
116.	The product of 39 and an even number will not be a Shola.
117.	There is a correspondence between odd integers and Sholas.
121.	Sholas are not closed with respect to subtraction.
128.	When two Sholas are added, the result is still divisible by 39.
134.	If x is a Shola, then $x^2 > 0$.
135.	If x is a Shola, then $-x^2 < 0$.

I.D.	Response
137.	Not all multiples of 39 are Sholas.
129.	Shola - Shola = even integer.

TEST ID

Write down as many mathematical statements as you can about the following function. Each statement should be such that it is true or would be true under certain conditions. Try to make each statement represent one idea only. Use your imagination.

$$y = 2^x, \; x \in R$$

I.D.	Response
1.	log y = x.
2.	When x is negative, $y = \frac{1}{2^{|x|}} = \frac{1}{2^{-x}}$.
3.	y cannot be a negative number.
3b.	y is always positive.
3c.	y can never be zero.
4.	2^x is an exponential function having a base of 2.
5.	Examples of ordered pairs: e.g., (2,4).
6.	y will always be greater than x for $x > 1$.
6a.	The value of y will always be greater than the value of x.
6b.	y will always be greater than x for $x \geq 0$.
7.	Algebraic rearrangements: e.g. $2^x - y = 0$
8.	y increases as x increases.
9.	
10.	If y cannot be divided by 2 evenly, then x is not an integer.
11.	
12.	The function has no minimum value.
13.	The function has no maximum value.
14.	The y intercept is 1.
14b.	There is no x intercept.
15.	There is only one intercept.

I.D.	Response
16.	The graph of the function is not symmetrical about the x or y axis.
17.	y can never be an odd number.
20.	If $x < 0$, then $0 < y < 1$.
21.	If $x > 0$, then $y > 1$.
22.	y is an integer if x is an integer.
23.	For every y in positive R there corresponds one and only one x.
24.	The lower x is, the closer to the x axis the graph of y goes.
25.	If y is negative, x has no <u>real</u> value.
26.	
27.	The graph is a wide partial parabola that never touches the x-axis.
28.	The graph is curved. It will never go to the third quadrant. It has no end.
29.	Domain: $x \in R$. Range: $y \in {}^{+}R$.
30.	This is a 1 - 1 function.
31.	x and y can never be equal.
32.	For every positive value of x (e.g., 0,1,2,3,4, etc.), y is doubled (e.g. 1,2,4,8,16, etc.).
33.	For every negative value of x(e.g., -1, -2, -3, etc) y is halved. (e.g. 1/2, 1/4, 1/8, etc.)

TEST 1B

Write down as many mathematical statements as you can about the following function. Each statement should be such that it is true or would be true under certain conditions. Try to make each statement represent one main idea only. Use your imagination.

$$y = 2^x(x^2-5x+6),\ x \in R.$$

I.D.	Response
1.	$y = 2^x(x-2)(x-3)$
2b.	Other rearrangements like 1.
2.	When x is negative, y is positive.
3.	When $x = 2$, $y = 0$.
4.	When $x = 3$, $y = 0$.
5.	$y = 0$, only if $x = 2$, or $x = 3$.
6.	
7.	The function has two x-intercepts, 2,3.
9.	Ordered pairs.
10.	The function is not linear.
11.	The y- intercept is 6.
12.	The graph of the function does not have symmetry about the x or y axis.
13.	Statements about the domain and range of the function.
2	y can be negative or positive depending on what x is.
26.	The function has an irregular graph.

TEST 2

Invent as many systems of equations as you can such that the solution set for each system is (5,7). Try to make each system as different from each other as possible. When you have thought out a pattern for making sequences, give two or three examples of the pattern, and group the similar systems together. Then look for a different pattern and group in a similar way. Please indicate the groups of systems that are different.

Note: This test was marked in terms of equations or systems of equations that included (5,7).

I.D.	Response
1.	Linear equations: e.g., $x+y = 12$, Form: $ax+by = c$.
2.	Quadratic equations: e.g., $x^2+xy+y^2 = 109$. Form: $ax^2 + by^2 + cxy = k$.
3.	Polynomial in two variables of order higher than 2.
4.	Equations with variable exponents: e.g., $x^y = 5^7$.

TEST II

Invent as many systems of equations as you can such that the solution set of each includes the numbers (1,2,3). Try to make the systems as different from each other as possible. When you have thought out a pattern for making sequences, give two or three examples of the pattern, and group the similar systems together. Then look for a different pattern and group in a similar way. Please indicate the groups of systems that are different.

I.D.	Response
1.	Equations of the first order: $ax+by+cz = k$ a,b,c, real numbers, x,y,z, variables.
2.	Equations of the second order: $ax^2 + by^2 + cz^2 + cxy + dxz + hyz =$ constant.
3.	Polynomial in three variables of order higher than 2.
4.	Equations with variable exponents.

TEST 2B

Think out and write down different sets of integers (m,n,q) such that

$$m^2 + n^2 = q^4.$$

Try to make the sets of triples as varied as possible. When you have found a pattern of triples, give two or three examples of the pattern, and group the similar triples together. Then look for a different pattern and group in a similar way. Please indicate the groups that are different.

I.D.	Response
1.	$(0,\ a^2,\ a)$ where a is an integer.
2.	$(a^2,\ 0,\ a)$ where a is an integer
3.	(0, 0, 0)
4.	(15, 20, 5)
5.	(24, 7, 5)

TEST IV

Show that if b and c are real, and x_1, x_2, are the roots of the quadratic equation $x^2 + bx + c = 0$, then $x_1 + x_2 = -b$, using as many distinct methods as you can. When you have thought out a method, write down an outline of the method, such that it would be possible to see how you would proceed if you had time. Then, look for another method. Try to think out and write outlines of as many methods as you can.

I.D	Response
1.	Using the completing the square method it is easy to show that $x_1 = (-b+ \ (b^2-4c))/2$, and $x_2 = (-b- \ (b^2-4c))/2$. Summing, we obtain $x_1 + x_2 = -b$.
1a.	Using specific real numbers for b and c, it is shown in as many cases as are attempted that the proposition is true.
2.	$x^2 + bx + c = (x-x_1)(x-x_2)$ Expanding and comparing coefficients, it follows that $x_1 + x_2 = -b$
3.	Substituting the roots in the equation and subtracting we obtain: $x^2_1 + bx_1 + c = 0 = x^2_2 + bx_2 + c$ Hence $x^2_, - x^2_2 + b(x_1 - x_2) = 0$. Assuming $x_1 \neq x_2$, the result follows after dividing.
4.	Graphing $y = x^2 + bx + c$ for several values of b and c. Observing the pattern, the result will be seen.
5.	By observing the symmetry of the parabola, it is clear that the sum of the roots of the quadratic equation is twice the x-value at the vertex. (Diagram furnished). This is

I.D. Response

5(cont'd) quite obvious when the parabola is symmetrical about the y-axis, in which case the sum of the roots is zero. The vertex is at the middle point of x_1 and x_2. Since the vertex is at $x = b/2$, the result follows.

TEST 5

Imagine that you wish to explain to Grade Six students why we cannot divide by zero. Think out some unusual mathematical approaches that you can use. List as many of these approaches as you can.

ID. Response

1. $a=b \rightarrow a^2-b^2 = ab - b^2$

$\rightarrow (a+b)(a-b) = b(a-b)$

$\rightarrow (a+b) = b$. If $a = 2$, $b = 2$,

$2+2 = 2 \quad 4 = 2$

Contradiction arises out of assuming that zero division is possible.

2. If $ab = c$ then $b = \frac{c}{a}$

$0 \times 6 = 0 \quad 6 = 0/0?$

$0 \times 54 = 0 \quad 54 = 0/0?$

$0 \times 17 = 0 \quad -17 = 0/0?$

$6 = 54 = -17 = 0/0$

False

3. $\frac{a}{a} = 1$ If $a = 0$ then

$\frac{0}{0} = 1 \quad 0 = 1$ False

4. If $a/b = c$, then $bc = a$.
But if $\frac{a}{0} = c$ then $0 \times c = 0$, not a. So it just don't work. Also $662 \div 0 = x$ but $x(0) \neq 662$, because any number multiplied by zero is equal to zero. $(x)0 = 0$.

ID.	Response
5.	$\frac{x}{0} \times \frac{3}{5} = \frac{3x}{0}$ $\frac{x}{0} \times \frac{3}{1} = \frac{3x}{0}$ $\frac{3}{1} \neq 3/5$
6.	When we divide a by b we are really subtracting b from a as often as possible. So if we try to take nothing (zero) away from something (any number 0), we'll have just as much as we had before so we could just keep going forever.
8.	Take a piece of candy, divide it into 2, then into 4 pieces. Take another piece and ask a student to divide it into zero pieces. There will always be at least one piece. No matter what we do we cannot make that piece disappear.
7.	For $x \neq 0$ $y \to 0$ $\frac{x}{y} \to \infty$ as y 0.
9.	Division by zero is not allowed because $\frac{x}{0}$ is not a real number, violating closure.
10.	$\frac{10}{x}$ as x decreases through positive values, tends to high POSITIVE values, as x approaches 0. $\frac{10}{x}$ as x <u>increases</u> from negative values tends to high NEGATIVE values. (Diagram supplied) This is a contradiction, if 10/0 is a number.
11.	Dividing a solid into n parts involves (n-1) cuttings. Dividing into 1 part involve 0 cutting. Dividing into 0 parts? -1 cutting?

TEST VI

Suppose that you are working in a system in which it is true that 2 ⊗ 8 = 4. Think out mathematical statements that would be true in this system, and in each case explain briefly (as far as you can) why.

I.D.	Response
1.	$a \otimes b = \frac{ab}{4}$
2.	$a \otimes b = a + b - 6$
3.	$a \otimes b = a - (b-10)$
4.	$a \otimes b = \frac{16a}{b}$
6.	$a \otimes b = a +$ (smallest non-one factor of b)
7.	$a \otimes b = a^{b/4}$
8.	$a \otimes b = a \, . \, b$
9.	$a \otimes b = b/a$
10.	$a \otimes b = ab(\text{mod } 12)$.
12.	For the system $a \otimes b = ab(\text{mod } 12)$ the identity element exists. $1 \otimes a = a$.
13.	For system $a \otimes b = b/a$, b can be any real number, but a cannot be zero since division by zero is undefined.
14.	$f(2 \otimes 8) = f(4)$
18.	$a \otimes b = 2a$
19.	$a \otimes b = b/2$
20.	$a \otimes b = a^4 = a^2$
22.	If a series of these operations were done it would matter in which order, i.e. $(a \otimes b \otimes c \neq a \otimes b \otimes c)$. [It seems logical and worksout in two test cases if different meanings of x].
23.	The operation could be defined in terms of some other standard operations. i.e. $+, -, \div$ x, etc.
26.	Examples of 8: $3 \otimes 27 = 9$; $4 \otimes 27 = 12$.
27.	$a \otimes b = 2a = b/2$.
28.	For system 27, $3 \otimes 12 = 6$.

I.D.	Response
30.	$2 \otimes 8 \neq 8 \otimes 2$.
31.	$a \otimes b = \sqrt{ab}$; $4 \otimes 16 = 8$.
32.	An extension of system 31 can be made such that $a \oplus b = 1/[2(a + b)]$. In this case $1/2 \oplus 1/2 = 1/2$.
33.	Following system 9 , $2 \otimes 10 = 5$; $3 \otimes 6 = 2$.
34.	For system 9, the larger a is for fixed b, the smaller the result will be.
45.	$n \otimes 4n = 2n$.
50.	$a \otimes a^2 = 2a$
51a.	$a \otimes b = b - a^2$.
52a	$a \otimes b = \dfrac{a^2 + b}{a + 1}$
59.	Using system 9, $a(a \otimes b) = a(\frac{b}{a} = b$.
60.	Assuming system 9($a \times b = b/a$.) Then $a \otimes (b \otimes c) = a^2 \otimes [(a \otimes b) \otimes c]$. Proof $a \otimes (b \otimes c) = c/ba$, and $(a \otimes b) \otimes c = ca/b$. Hence $a^2 \otimes [(a \otimes b) \otimes c] = a^2 \otimes ca/b = c/ba$ $= a \otimes (b \otimes c.)$
61.	Assuming system 9, $\dfrac{1}{b \otimes a} = a \otimes b$, since $b/a = 1/(a/b)$.
62.	Assuming 61, $(a \otimes b)(b \otimes a) = 1$.
63.	Assuming systems 9, $\log (a \otimes b) = \log b - \log a$, since $\log (\frac{b}{a})$ $= \log b - \log a$.
64.	Assuming system 9, $(a \otimes b)^x = a^x \otimes b^x$
70a.	$a \otimes b = b - 2a$.
84.	$a \otimes 0$ will always be equal to 0 except when $a = 0$ (Assuming system 9)

ID.	Response
89.	Following 2 ⓧ 8 = 4, define 8 ⓓ 4 = 2 Hence 10 ⓓ 2 = 5 because 5 x 2 = 10.
94.	Assuming system 10, the system mod 12, the cummutative property holds a ⓧ b = b ⓧ a.
96.	Assuming system 10, the x inverse element would be any number such that a ⓧ b = 1, where be is any number which would make ab to be <u>one</u> larger than 0, 12, 24, 36, etc.
99.	Assuming system 9. If the elements of the system are all integers, then the b element must be a multiple of the a element.
101.	a ⓧ b = $(\sqrt[3]{b}\ )^a$, and b, integers.
102.	a ⓧ b = c, where c is the greatest even number less than or equal to (a + b)/a.
105a.	a ⓧ b = 1 + $(\log_a b)$.
106a.	a ⓧ b = b - 2a.

TEST 6B

T(n) is defined as the set of all positive integers less than n which divide n evenly. Think out and guess as many properties of T(n) as you can.

ID.	Response
1.	T(n) = $\emptyset$, if n = 1. [Note: $\emptyset$ indicates the empty set]
2.	T(n) = $\{1\}$ if n $\neq$ 1 and is prime.
3.	If n >1 and is even, then the greatest member of T(n) is n/2.
4.	For all n, zero is <u>not</u> a member of T(n). Also, n > 0.
5.	The greatest member of T(n) is not greater than n/2.

ID	Response
6.	For any n, multiplication of members within T(n) will not always give members of the set. The product may exceed n.
6b.	T(n) is not closed with respect to addition.
6c.	T(n) is not closed with respect to division.
6d.	T(n) is not closed with respect to subtraction.
7.	The natural number 1, is a member of T(n) unless n = 1.
8.	The sum of all members of T(n) is less than 2n.
9.	If x is a member of T(n), then $0 \leq x \leq n$.
10.	If n is odd, then T(n) consists entirely of odd numbers.
10b.	If n is even, then T(n) consists of odd and even integers <u>OR</u> entirely of even integers.
11.	Specific Examples: e.g. $T(6) = \{1, 2, 3\}$
13.	If a x b = n, then both a and b are members of T(n).
15.	If n is even, then 2 is an element in T(n).
16.	T is not a function, since for every n, there are at least 2. T(4) = 1, 2.
17.	The set is finite.
18.	The members may be even or odd.
22.	The number of members of T(n) is less than n.
23.	The number of members of T(n) is odd if n is not a perfect square.
25.	Sum of members of T(n) = n, if n is a perfect number.
26.	If $n = a^b$, where a is a prime, then $T(n) = \{a, a^2, a, \ldots a^{b-1}\}$

I.D.	Response
28.	If n = 12!, then it will be evenly divisible by (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12). It will also be divisible by 2^{10}, 3^{5} , 4^{5}).
29.	If n = 12!, there will be six prime numbers in T(n).
30.	If n = 12!, all members of T(n), will be divisible by (1, 2, 3, 5, 7, 11).

APPENDIX D

TESTS (ORIGINAL DRAFT)

321 Education Building

January 14, 1970

Dear Dr.

I am enclosing tests which I intend to try out next week. I shall be grateful for your help and comment on their face and content validities.

I shall be grateful if you will give a rating to each test with regard to the ability as defined that I am proposing to test for. I shall be grateful if you will indicate the rating using symbols as follows:

Rating Symbol	Meaning
V	A very good test of the ability.
G	A good test of the ability.
S	A satisfactory test of the ability.
U	Unsatisfactory and unsuitable.

In addition, I shall be grateful if you will also rank the tests for each ability in order of suitability, using natural numbers 1, 2, 3, ..., where 1 indicates the most suitable.

I shall be grateful to have your general comments.

Thank you very much indeed for your help.

Yours sincerely

J. Modupe Taylor-Pearce

DIVERGENT PRODUCTION OF MATHEMATICAL UNITS

Ten minutes will be allowed for each test. Open Book.

DEFINITION: The ability to produce various elementary mathematical ideas related to a mathematical situation.

Test No.	Test	Rating	Rank
1.	Write down ten mathematically true statements about an Epudom in the sense defined below: An Epudom is an even integer divisible by 35.		
2.	Write down ten mathematically true statements about a kasep in the sense defined below: A kasep is an odd integer divisible by 39.		
3-8.	A mathematical function will be presented in tests 3-8. Students are requested to think out and write down several mathematical statements about the function. Each statement should be such that it is true or would be true under certain conditions. Try to make each statement represent one main idea only. Use your imagination.		
3.	$y = x^2-7x+10, \quad x \in R.$		
4.	$y = 2^x, \quad x \in R.$		

Test No.	Test	Rating	Rank
5.	$y = 2^x(x^2-5x+6), \quad x \in R.$		
6.	$y = 2^x \log x, \quad x \in R.$		
7.	$y = x(\log x), \quad x \in R.$		
8.	$y = (\log x)^2 - 5(\log x) + 6, \quad x \in R.$		
9.	Give five examples of relations which are not transitive. (symmetric? reflexive?)		
10.	Give five examples of operations which are not commutative. (associative? no-inverse?)		

DIVERGENT PRODUCTION OF MATHEMATICAL CLASSES

Ten minutes will be allowed for each test. Open Book.

DEFINITION: The ability to resist fixedness in mathematical thinking and to produce mathematical ideas that are different, in relation to a mathematical situation.

Test No.	Test	Rating	Rank
1.	Think out and write down different sets of integers (m, n, q) such that $m^2+n^2 = q^2$. Try to make each set of triples as different from the others as possible.		
2.	Invent several different systems of equations such that the solution set for each is (5,7). Try to make the systems of equations as different from each other as possible.		
3.	Think out and write down <u>different</u> sets of integers (m, n, q) such that $m^2+n^2=q^4$. Try to make each set of triples as different from the others as possible.		
4.	Invent several <u>different</u> systems of equations such that the solution set for each is (1,2,3).		

Test No.	Test	Rating	Rank
5.	Invent five <u>different</u> equations that have no real number solution.		
6.	Invent five <u>different</u> equations that have real but not rational number solutions.		
7.	Show that $\frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2}$, using up to five different methods.		
8.	Show that if b and c are real, and x_1, x_2 are the roots of the quadratic equation $x^2+bx+c = 0$, then $x_1+x_2 = -b$, using up to five different methods.		
9.	Prove any mathematically true statement (or theorem) that you know in up to five different ways.		

DIVERGENT PRODUCTION OF MATHEMATICAL RELATIONS

Ten minutes will be allowed for each test. Open Book.

DEFINITION: The ability to produce or recognize mathematical relationships.

Test No.	Test	Rating	Rank
1.	The following numbers in the form of an infinite series represent a definite pattern. Please think out various patterns that could account for the numbers as they are arranged. In each case state the value of x and y, and explain briefly the indicated pattern. $\log_2 4$, $\log_2 16$, 2^3, x, y, ...		
2.	A student made the following false mathematical statement: $\log_2 4 = 16$. Assuming that he is likely to make the same type of error in making further statements of this type, what possible numbers could he write down for $\log_3 5$? Write down five possible answers, and in each case explain the systematic reasoning which would account for your answer.		

Test No.	Test	Rating	Rank
3.	Write down several general relations on the set of real numbers which could include the following ordered pairs: [(-2,4), (0,0), (2,4)].		
4.	Given A = (10, 13, 23, 36, 59), invent several binary relations that could be defined on A and in each case list the ordered pairs.		
5.	The following three functions are arranged in a definite pattern, and are part of an infinite series. Think out five possible functions that could stand in place of f(x), and in each case explain briefly how you obtained the function. (x^2+2x+1), (x^2+6x+9), $f(x)$, ...		
6.	Given that Z = ($\sqrt{3}$, 3^3, $\log_3 3$, 1/3, $3^{1}/_{3}$) invent several binary relations that could be defined on Z, and in each case list the ordered pairs.		

DIVERGENT PRODUCTION OF MATHEMATICAL SYSTEMS

Ten minutes for each test. Closed Book.

DEFINITION: The ability to organize elementary mathematical ideas into complex ideas.

Test No.	Test	Rating	Rank
1.	The student will be given back the answer sheet containing his responses to the test on Divergent Production of Mathematical Units. For each set of test responses the student will be given the following test: Invent five mathematical problems using some or all of the ideas you have listed in your response sheet to test (?). Each problem should involve at least two of the ideas you listed. You are free to use ideas from other areas.		
2.	Make five different groupings of the ideas you have listed in your response sheet to test (?). You may have as many sets as you wish in each grouping. Please explain briefly the reason for (or relation in) each grouping.		

DIVERGENT PRODUCTION OF MATHEMATICAL TRANSFORMATIONS

Ten minutes for each test. Closed Book.

DEFINITION: The ability to produce original responses involving reinterpretations and redefinitions.

Test No.	Test	Rating	Rank
1.	Make up five three choices multiple choice questions involving quadratic equations. Try to make your questions as original and as unusual as you can.		
2.	Make up five three choices multiple choice questions involving relations and functions. Try to make your questions as original and as unusual as you can.		
3.	Imagine that you wish to explain to various Grade Six students why we cannot divide by Zero. Think out some unusual mathematical approaches that you can use. List five of these approaches.		
4.	Imagine that you wish to explain to various Grade Five students why it is true that $-1 \times -1 = 1$. List five different approaches you might use.		

Test No.	Test	Rating	Rank
5.	Invent <u>five</u> different operations on numbers each of which behaves in an unusual way. In each case define the operation carefully.		
6.	Invent <u>five</u> different functions which behave in unusual ways. In each case define each function carefully.		

DIVERGENT PRODUCTION OF IMPLICATIONS

Ten minutes. Open Book.

Note: The content for testing the ability has been deliberately made to include material with which the students should normally be unfamiliar.

DEFINITION: The ability to produce mathematical implications from a given set of conditions.

Test No.	Test	Rating	Rank
1.	Suppose that you are working in a system in which for real numbers a, n, (i) $a^{n+1} = a^n \times a$. (ii) $a^{12} = a$. Think out and guess ten mathematical statements such that the above two conditions would be true for each of them.		
2.	Suppose that you are working in a system in which it is true that $2 \otimes 8 = 4$. Think out and guess ten other statements that would be true in this system, and explain briefly (as far as you can) why.		

Test No.	Test	Rating	Rank
3.	T(n) is defined as the set of all positive integers less than n which divide n evenly. Think out and guess as many properties of T(n) as you can.		
4.	P(n) is defined as the set of all positive integers less than n^n which n divides evenly. Think out and guess as many properties of P(n) as you can.		
5.	Suppose you are working in a system in which the following condition is true: Given a line L and a point P not on L, there are at least two lines L_1, L_2, which contain P and are parallel to L. Think out and guess <u>five</u> implications of this condition.		

Test No.	Test	Rating	Rank
6.	Let A = (a,b,c,d) and let x, y, z, stand for ANY of the elements of A. Let Θ be an operation in A such that the following properties are true: 1. x Θ y is a unique element in A. 2. x Θ (y Θ z) = (x Θ y) Θ z. 3. There exists a unique element e of A, such that x Θ e = e Θ x = x. 4. For every x in A, there exists a unique element x in A such that x Θ x = x Θ x = e. Think out and guess ten true statements about A or elements in A, using the four properties. (20 minutes).		

GOAL DIRECTED SYNTHESIS

Twenty minutes for each problem. Open Book.

DEFINITION: The ability to solve problems which demand ingenuity.

Test No.	Test	Rating	Rank
1.	Find all two digit natural numbers, x, such that the product of the digits equals $x^2-10x-10$.		
2.	Find all two digit natural numbers, y, such that the sum of the digits of y is $y^2-20y+9$.		
3.	Find all positive integers x, such that $x(x+3)$ is the square of an integer.		
4.	Prove that the product of two consecutive integers can never be the square of an integer.		
5.	Prove that there are no positive integers, y, such that $y(y+4)$ is the square of an integer.		
6.	Prove that there are no integral solutions for the equation $x^2-2px-2q = 0$, where p and q are odd integers.		
7.	Prove that there are no rational solutions for the equation $x^2-2px-2q = 0$, where p and q are integers.		

TABLE 1

SUMMARY OF ANALYSIS OF RATINGS OF JUDGES ON APPROPRIATENESS OF DP TESTS[1]

Ability	Test No.	N(V)	N(G)	N(S)	N(U)	$\frac{N(V,G,S)}{N(V,G,S,U)}$	AR	I(AR)
DMaU	1	4	2	0	0	1.00	2.67	G-V
	2	4	2	0	0	1.00	2.67	G-V
	3	2	3	1	0	1.00	2.17	G-V
	4	2	3	1	0	1.00	2.17	G-V
	5	2	3	1	0	1.00	2.17	G-V
	6	2	3	1	0	1.00	2.17	G-V
	7	2	3	1	0	1.00	2.17	G-V
	8	1	4	1	0	1.00	2.00	G
	9	1	3	0	2	0.67	1.50	S-G
	10	1	3	0	2	0.67	1.50	S-G
DMaC	1	2	2	1	1	0.83	1.80	S-G
	2	5	1	0	0	1.00	2.80	G-V
	3	2	2	2	0	1.00	2.00	G
	4	3	2	1	0	1.00	2.30	G-V
	5	1	3	2	0	1.00	1.80	S-G
	6	1	3	2	0	1.00	1.80	S-G
	7	3	1	1	1	0.83	2.00	G
	8	2	1	2	1	0.83	1.70	S-G
	9	2	2	0	2	0.67	1.70	S-G

1

N(V) denotes the number of judges who rated the test as V.
N(G) denotes the number of judges who rated the test as G.
N(S) denotes the number of judges who rated the test as S.
N(U) denotes the number of judges who rated the test as U.
N(V,G,S) denotes the number of judges who rated the test as at least S.
N(V,G,S,U) denotes the total number of judges
AR denotes the average rating of the judges when the weights 3,2,1,0, were assigned to V,G,S,U, respectively.
I(AR) denotes the interpretation of the average rating.

TABLE 1 (CONTINUED)

SUMMARY OF ANALYSIS OF RATINGS OF JUDGES ON APPROPRIATENESS OF DP TESTS[1]

Ability	Test No.	N(V)	N(G)	N(S)	N(U)	$\frac{N(V,G,S)}{N(V,G,S,U)}$	AR	I(AR)
DMaR	1	2	2	1	1	0.83	1.80	S-G
	2	3	2	0	1	0.83	2.20	G-V
	3	3	2	1	0	1.00	2.30	G-V
	4	2	4	0	0	1.00	2.30	G-V
	5	2	3	1	0	1.00	2.20	G-V
	6	1	3	2	0	1.00	1.80	S-G
DMaS	1	0	1	0	3	0.50	0.67	U-S
	2	0	2	2	2	0.67	1.00	S
DMaT	1	1	1	2	2	0.67	1.20	S-G
	2	1	1	2	2	0.67	1.20	S-G
	3	1	3	1	1	0.83	1.70	S-G
	4	1	3	1	1	0.83	1.70	S-G
	5	2	1	2	1	0.83	1.70	S-G
	6	1	2	3	0	1.00	1.70	S-G
DMaI	1	2	2	2	0	1.00	2.00	G
	2	3	2	0	1	0.83	2.20	G-V
	3	3	1	1	1	0.83	2.00	G
	4	3	1	1	1	0.83	2.00	G
	5	3	0	3	0	1.00	2.00	G
	6	1	1	3	1	0.83	1.3	S-G
GDS	1	3	3	0	0	1.00	2.5	G-V
	2	3	3	0	0	1.00	2.5	G-V
	3	3	3	0	0	1.00	2.5	G-V
	4	2	3	1	0	1.00	2.17	G-V
	5	1	4	1	0	1.00	2.00	G
	6	2	3	1	0	1.00	2.17	G-V
	7	2	3	1	0	1.00	2.17	G-V

[1] See footnote of Table 2 above.

TABLE 2

TESTS SELECTED AS A RESULT OF CONTENT VALIDATION

CATEGORY	TEST NUMBERS[1]
DMaU	1, 2, 3, and 4
DMaC	2, 3, and 4
DMaR	2, 3, 4, and 5
DMaS	None[2]
DMaT	3, 4, 5, and 6
DMaI	1, 2, 3, and 4
CPS (Goal Directed Synthesis)	1, 2, 3, and 4

[1]The test numbers correspond to the numbers of the tests as found in this appendix.

[2]Tests DMac 1, 5, and 6 were later adapted to serve as DMaS tests.